T.Y. B.SC. APPLIED STATISTICS AND DATA ANALYTICS (HONS.)
ACADEMIC YEAR 2023-24

REGIONAL AND STATE-WISE ELECTRICITY CONSUMPTION FORECASTING IN INDIA

ANALYSIS OF TIME SERIES

Submitted By:
Ayush Anand D003
Senorita Chettiar D004
Ayush Pujari D011
D Vaishnavi D021

Submitted To:
Mr. Priyesh D. Tiwari
Assistant Professor
Department of Statistics
Mithibai College (Mumbai)

PROJECT OVERVIEW

This project aims to forecast electricity consumption across different regions in India using historical data from January 2013 to December 2023. The study utilizes data sourced from the Power System Operation Corporation Limited (POSOCO) to analyze consumption patterns and predict future demand. The analysis involves aggregating data into four major regions — North, South, East, and West — and employing two prominent time series forecasting methods: Holt-Winters Exponential Smoothing and Seasonal ARIMA (SARIMA). In addition to regional analysis, we also perform a focused analysis on the top three states with the highest consumption. The best-performing models are used to forecast electricity consumption for 2024, offering insights for energy management and policy planning.

OBJECTIVES

To forecast regional electricity consumption across North, South, East, and West regions of India for the year 2024 using historical data.

To conduct individual time series analysis for the top three states with the highest consumption to understand localized consumption patterns.

Dataset Overview

Source

The dataset for this project was sourced from the Power System Operation Corporation Limited (POSOCO) website, a reliable source for electricity load and consumption data in India.

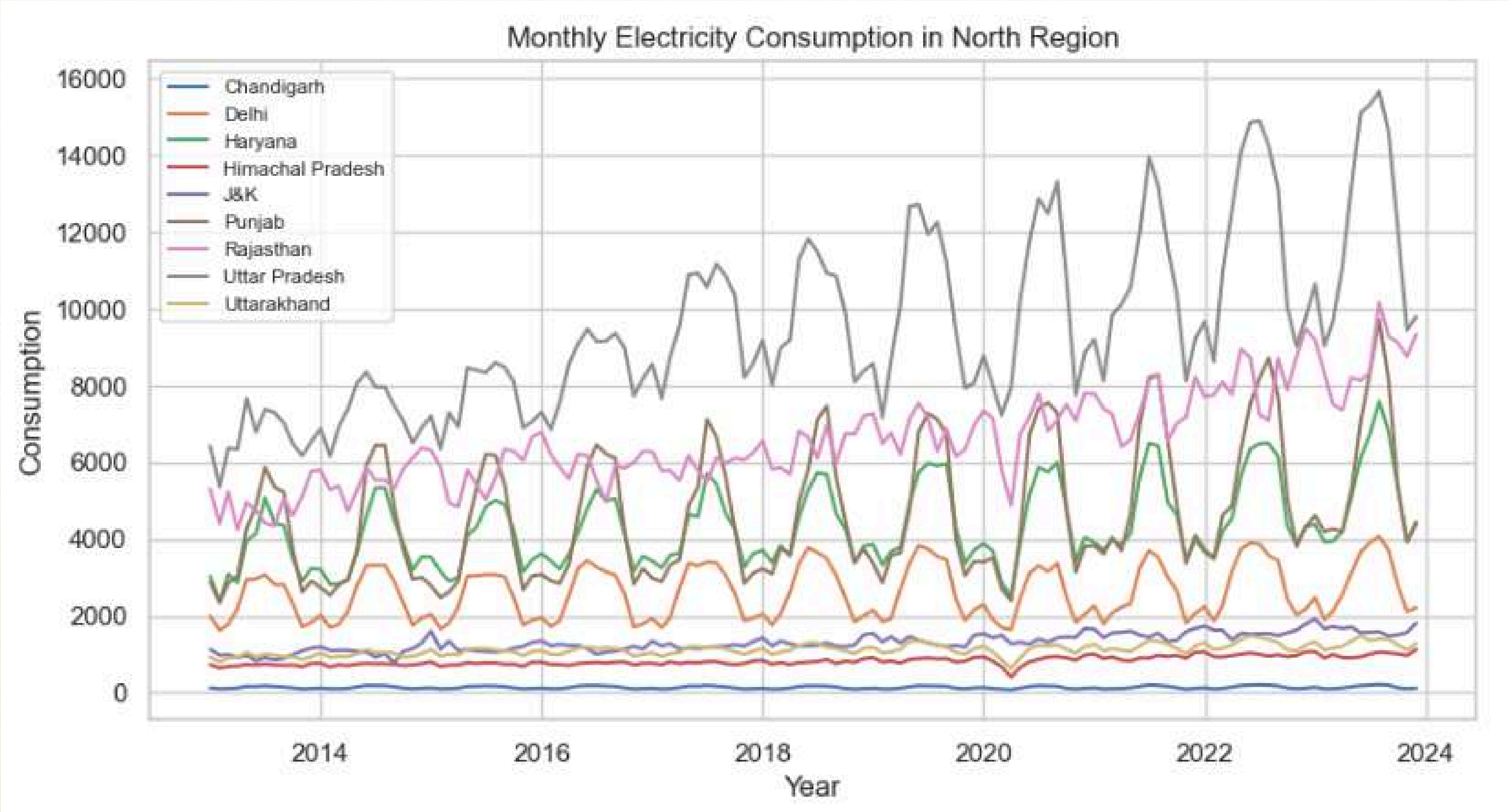
Key Features

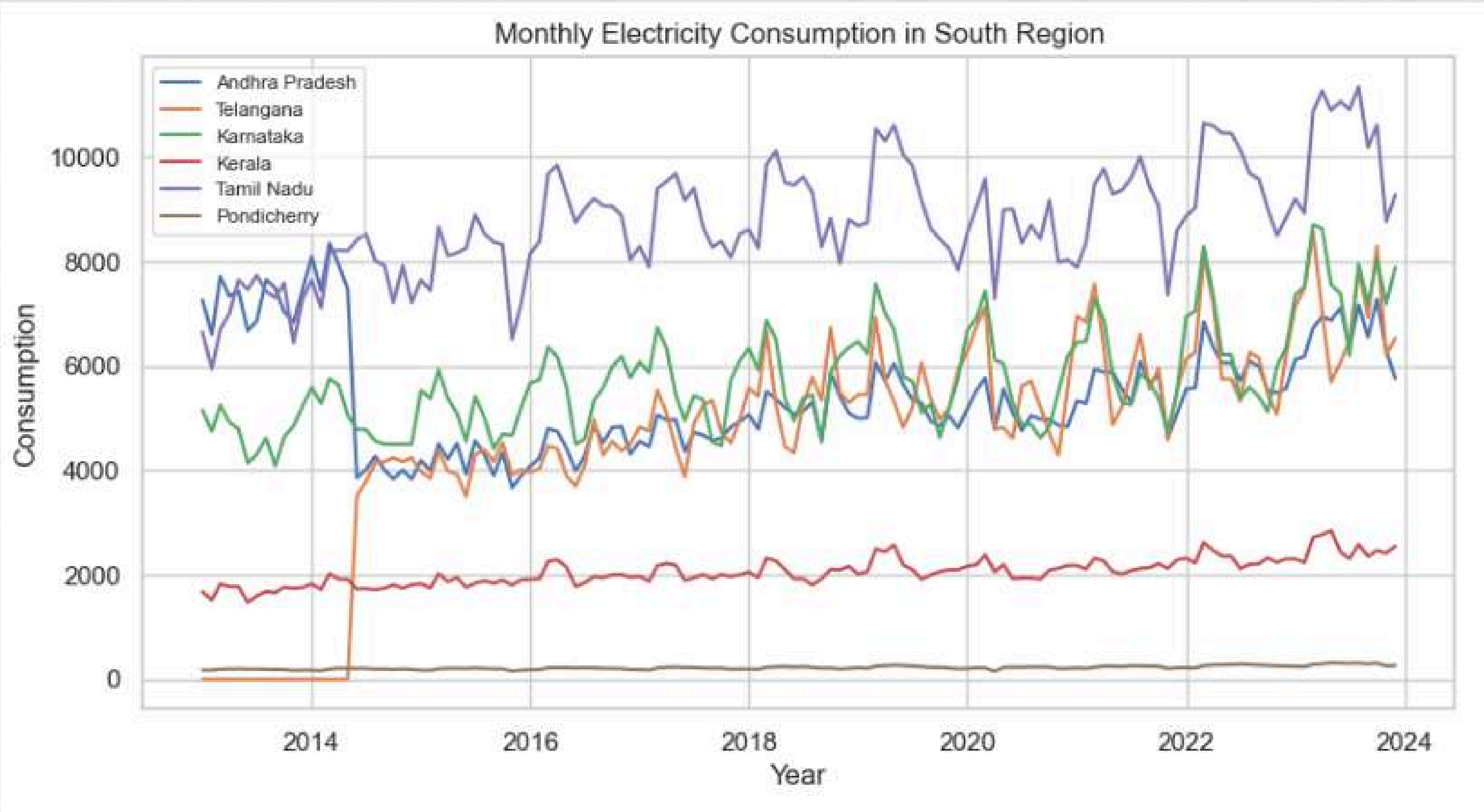
- Time Period: January 2013 - December 2023 (11 years of data)
- Frequency: Monthly
- Geographical Scope: Covers all states and union territories in India
- Variables:
  - Date: The month and year of the consumption data
  - Electricity Consumption: Measured in million units (MU) for each state and union territory

Regional Aggregation

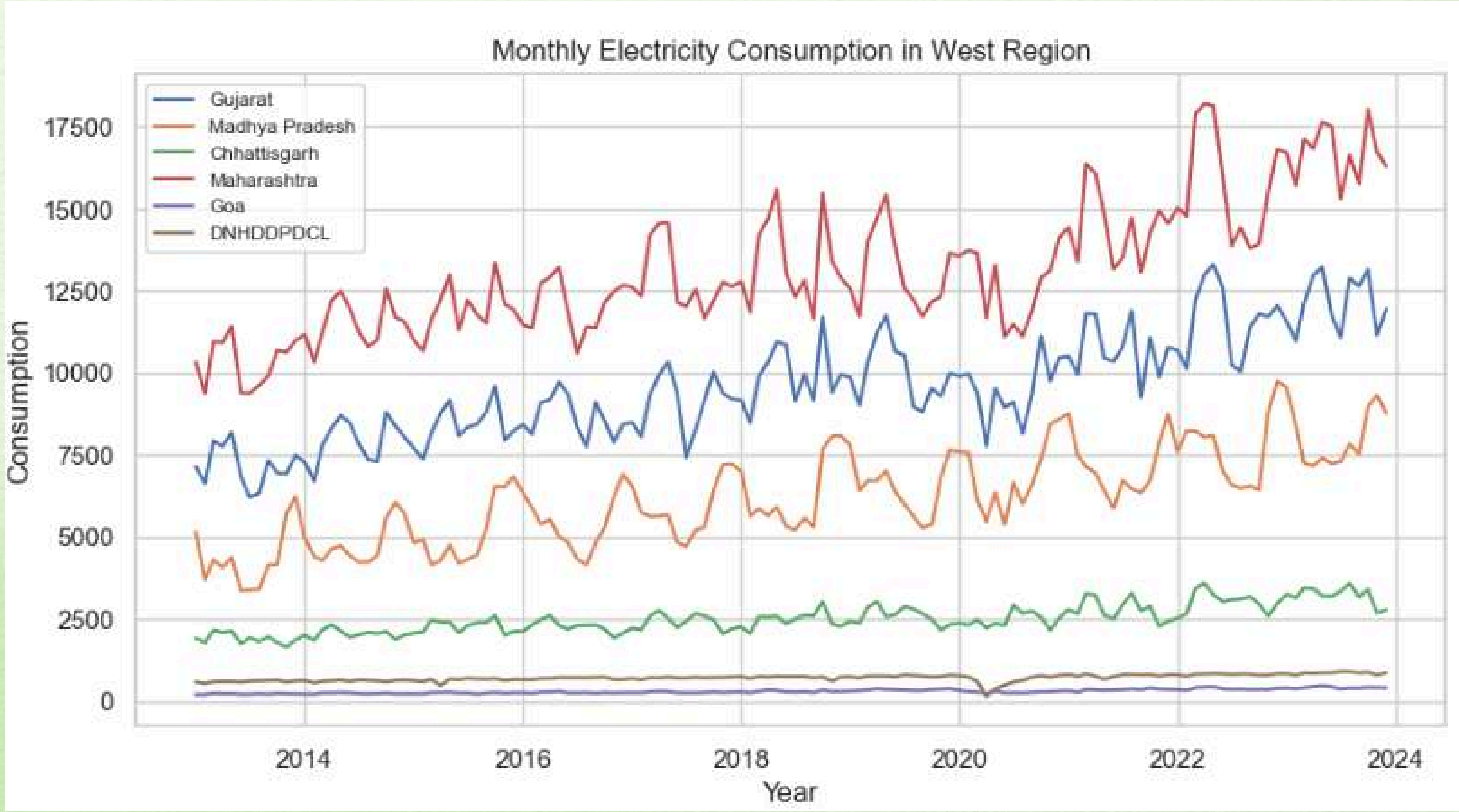
The state-wise data has been combined to create four aggregated datasets for the North, South, East, and West regions of India. This regional aggregation helps in capturing broader consumption patterns and trends across different parts of the country.

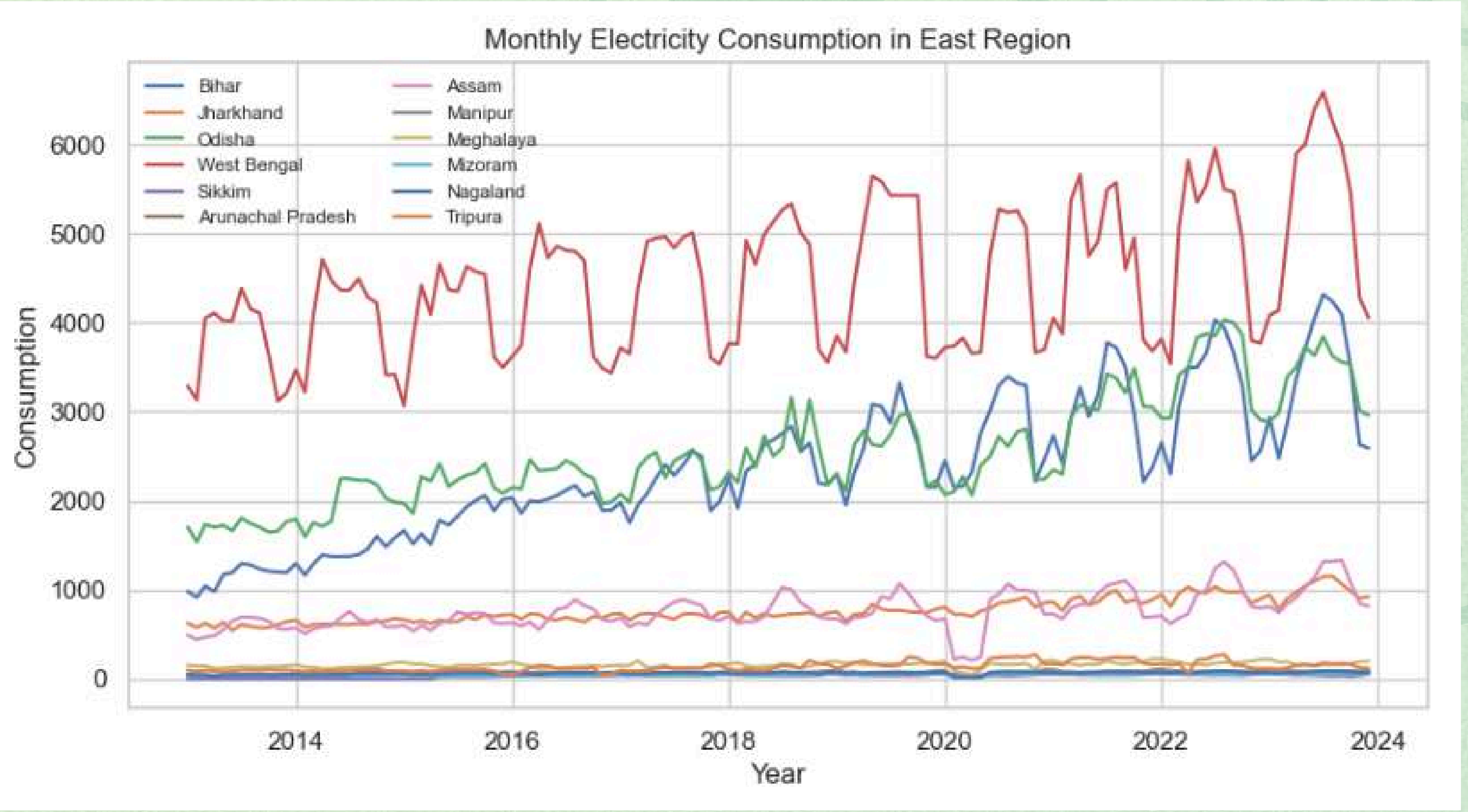
State-Wise Electricity Consumption Trends in Northern and Southern Regions





State-Wise Electricity Consumption Trends in Western and Eastern Regions





Overview of Forecasting Approach

01.

Data Analysis Insights: Observed distinct trends and seasonal patterns in electricity consumption data across states.

02.

Modeling Techniques: Applied SARIMA and Holt-Winters models to capture these patterns effectively.

03.

Validation Process: Conducted a train-test split to evaluate model performance.

Train data:
Jan 2013-Dec 2022

Test data:
Jan 2023-Dec 2023

04.

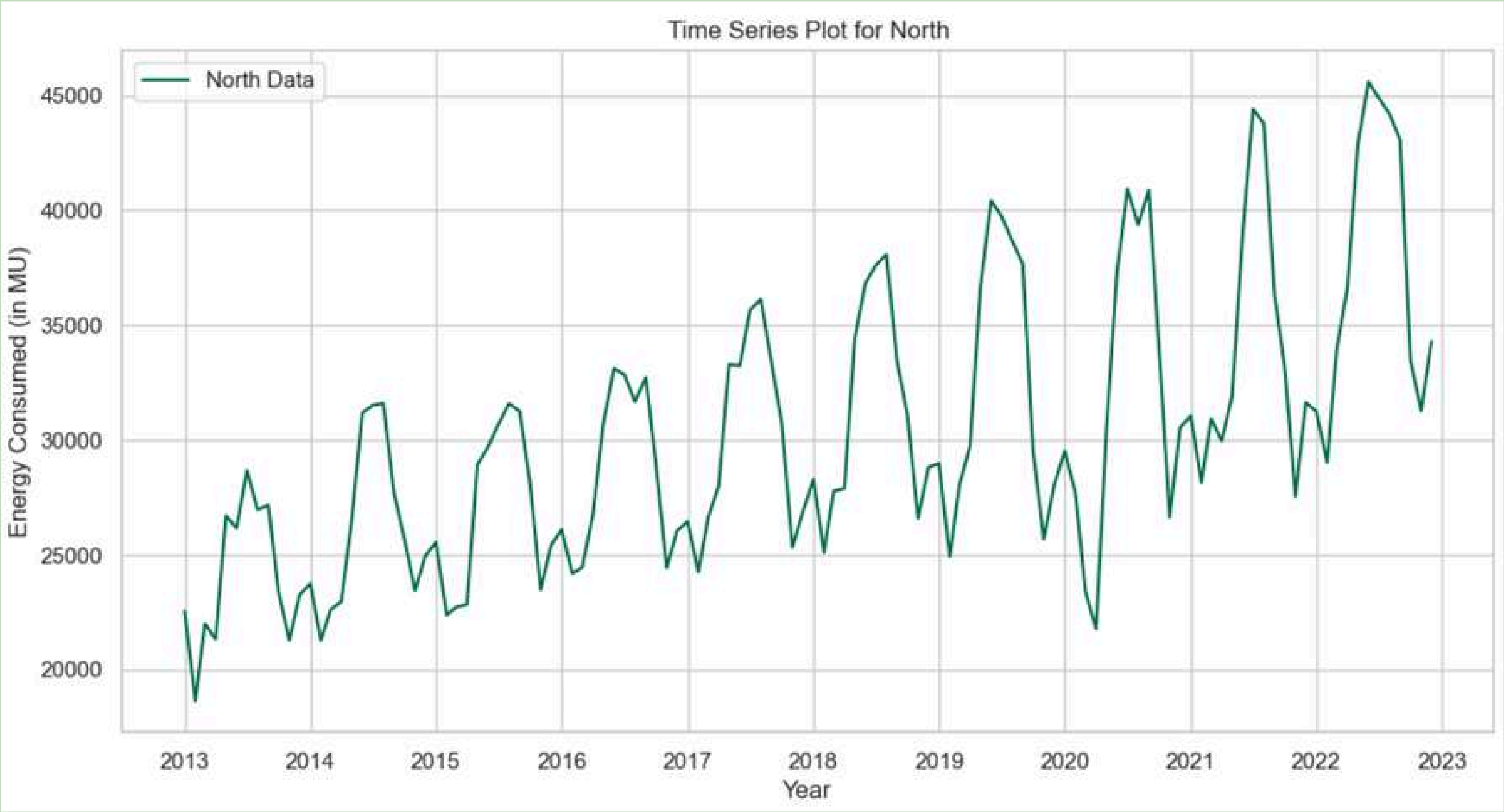
Model Comparison: Compared the accuracy of SARIMA and Holt-Winters models to select the best fit for each region.

05.

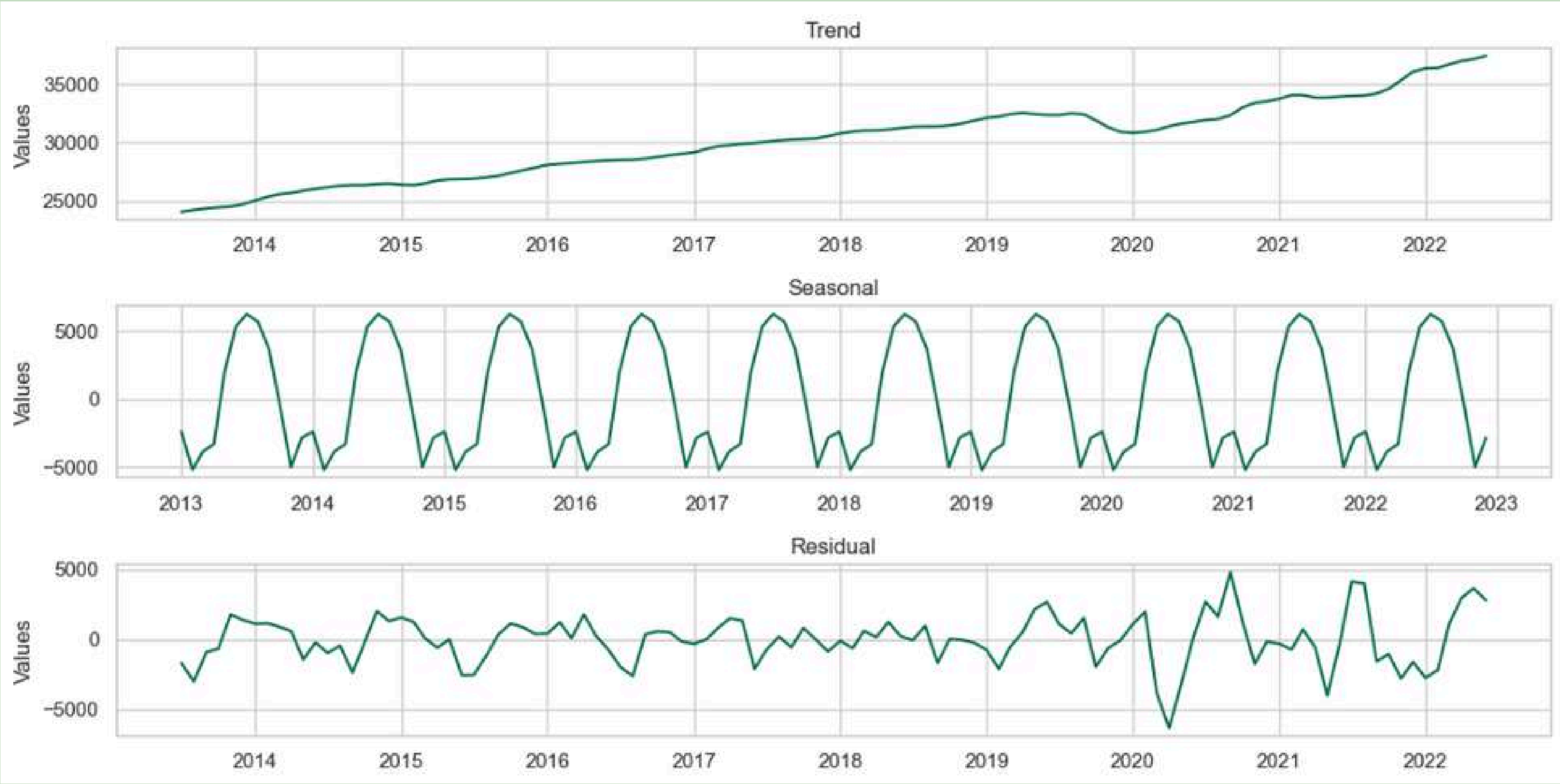
Outcome: Ensured robust and tailored forecasts for each region based on the selected models.

Electricity Consumption Analysis for North Region

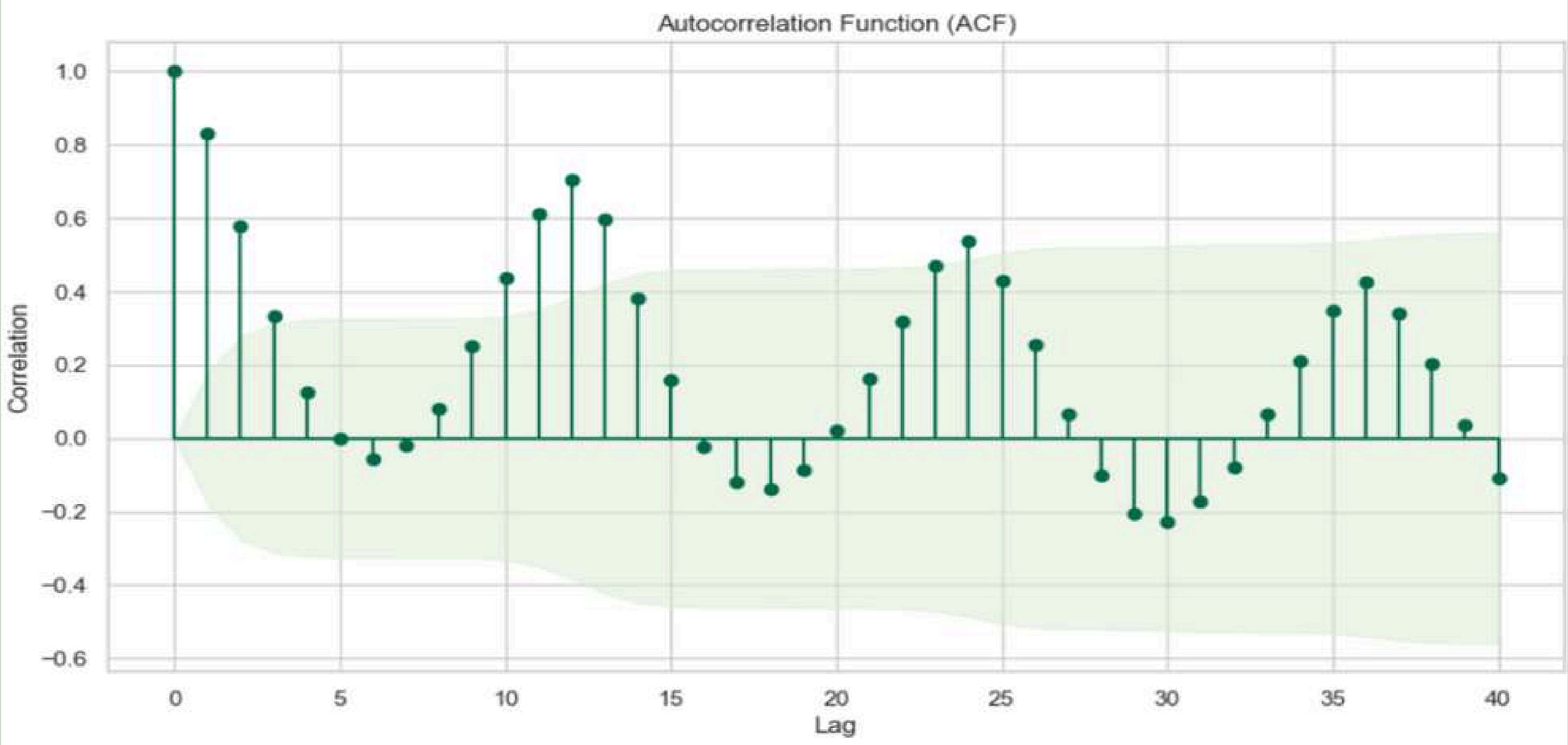
Time Series Plot



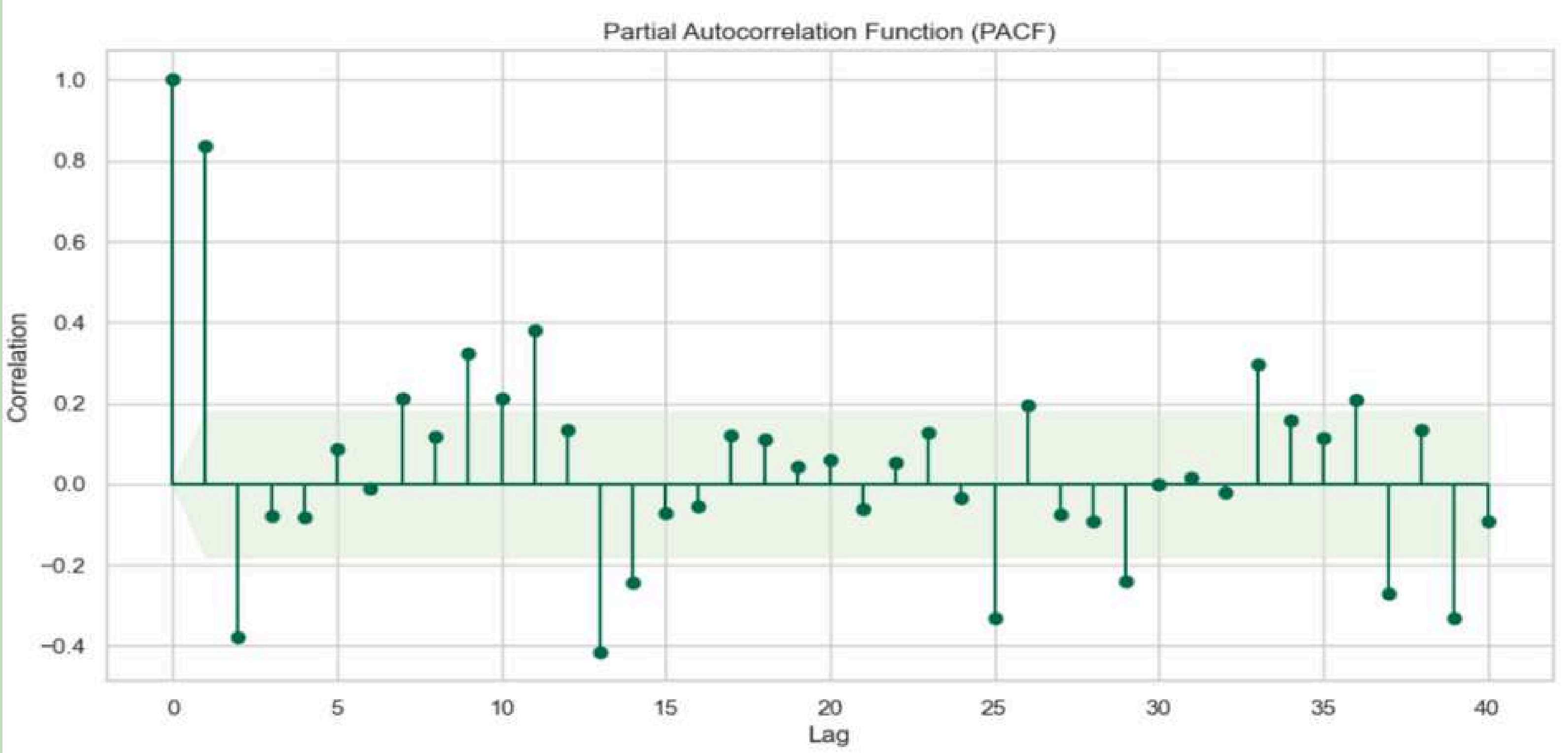
Time Series Decomposition Plot



Autocorrelation Function (ACF) Plot



Partial Autocorrelation Function (PACF) Plot



Stationarity Check

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(train_data)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: 0.536684
p-value: 0.985946
```

- Positive ADF Statistic
- A p-value of 0.985946 (which is greater than L.o.S 0.05)

Thus, we fail to reject the Null Hypothesis of the data being non-stationary and can safely conclude that our data is non-stationary.

Also, from the decomposition performed, we can also conclude that seasonality is present in the data at

Preprocessing for Stationarity: Seasonal and Normal Differencing

After Seasonal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(sd_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -2.302620
p-value: 0.171124
```

After Normal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(double_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -4.526441
p-value: 0.000176
```

The time series data is stationary after applying seasonal and normal differencing, as indicated by the p-value being less than 0.05 from the ADF test.

Holt-Winters Linear Exponential Smoothing Model Results

Holt-Winters Model Summary

Dep. Variable:	North	**No. Observations:**	120
Model:	ExponentialSmoothing	**SSE**	460066439.190
Optimized:	True	**AIC**	1851.127
Trend:	Additive	**BIC**	1895.727
Seasonal:	Additive	**AICC**	1857.899
Seasonal Periods:	12	**Date:**	Wed, 04 Sep 2024
Box-Cox:	False	**Time:**	04:05:48
Box-Cox Coeff.:	None		

	coeff	code
smoothing_level	1e-05	alpha
smoothing_trend	2e-05	beta
smoothing_seasonal	9e-06	gamma

Holt-Winters Model Equations

Holt-Winters Model

The Holt-Winters model equations based on the provided model summary are as follows:

Level Equation

$$L_t = 0.000001(Y_t - S_t) + (1 - 0.000001)(L_{t-1} + T_{t-1})$$

Trend Equation

$$T_t = 0.00002(L_t - L_{t-1}) + (1 - 0.00002)T_{t-1}$$

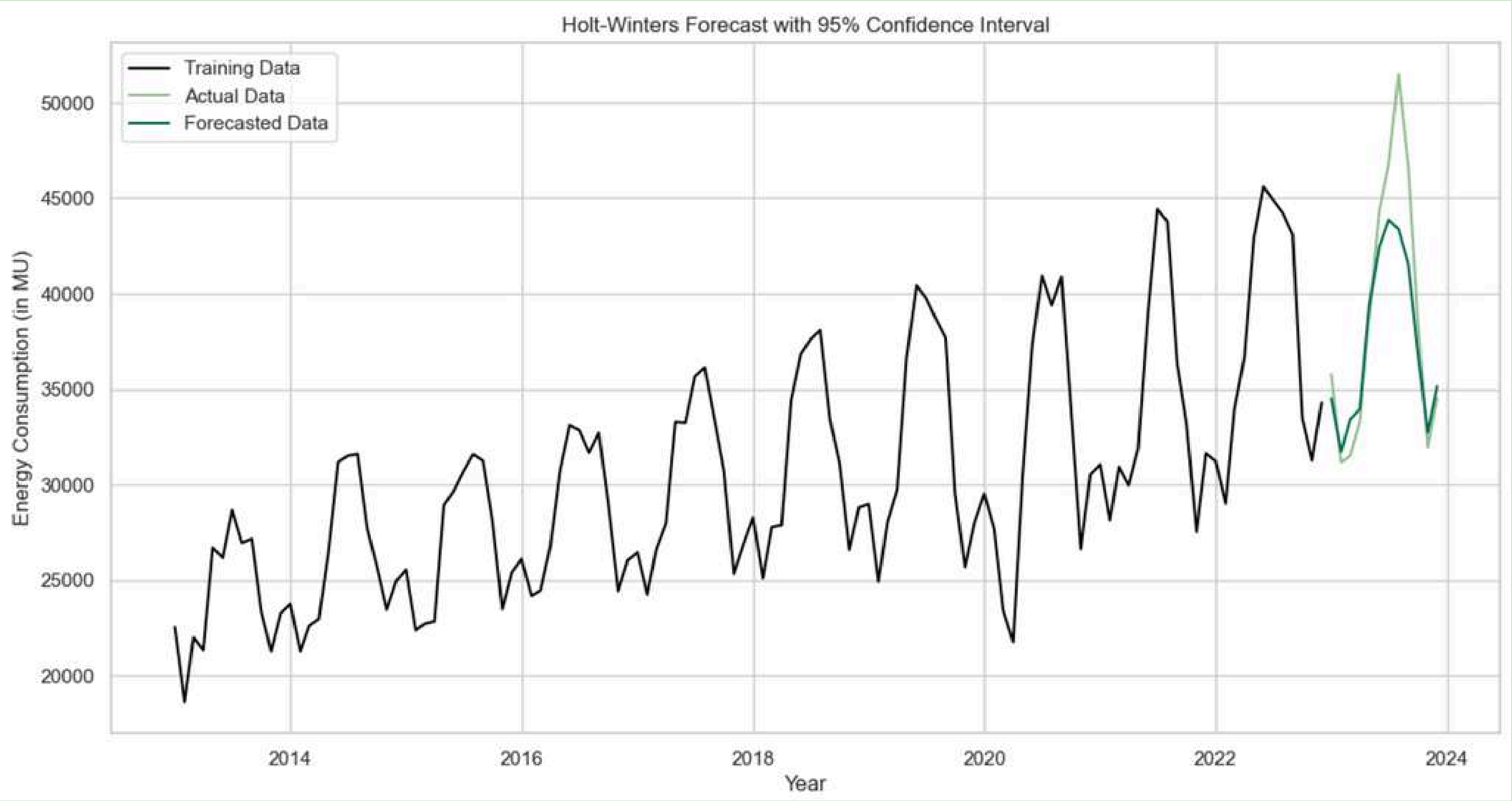
Seasonal Component

$$S_t = 0.000009(Y_t - L_t) + (1 - 0.000009)S_{t-12}$$

Forecast Equation

$$\hat{Y}_{t+h} = L_t + h \cdot T_t + S_{t+12-h}$$

Holt-Winters Model Fit and Forecasts



Holt-Winters Model Accuracy Metrics

MAPE (Mean Absolute Percentage Error) → **5.01%**

MSE (Mean Squared Error) → **9490577.1972927**

SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Summary

Dep. Variable:	North	No. Observations:	120
Model:	SARIMAX(2, 1, 0)x(2, 1, [1, 2], 12)	Log Likelihood	-732.107
Date:	Wed, 04 Sep 2024	AIC	1478.215
Time:	09:29:41	BIC	1494.976
Sample:	01-01-2013	HQIC	1484.940
	- 12-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0715	0.092	-0.775	0.439	-0.252	0.109
ar.L2	-0.2650	0.102	-2.600	0.009	-0.465	-0.065
ar.S.L12	0.2371	0.194	1.221	0.222	-0.144	0.618
ar.S.L24	-0.5048	0.147	-3.442	0.001	-0.792	-0.217
ma.S.L12	-1.0601	0.219	-4.831	0.000	-1.490	-0.630
ma.S.L24	0.6221	0.184	3.375	0.001	0.261	0.983
sigma2	4.046e+06	6.74e+05	6.005	0.000	2.73e+06	5.37e+06

Ljung-Box (L1) (Q):	0.23	Jarque-Bera (JB):	1.76
Prob(Q):	0.63	Prob(JB):	0.42
Heteroskedasticity (H):	2.86	Skew:	-0.36
Prob(H) (two-sided):	0.01	Kurtosis:	3.10

SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\mathrm{SARIMA}[2, 1, 0][2, 1, 1, 12]$$

Forecast Equation

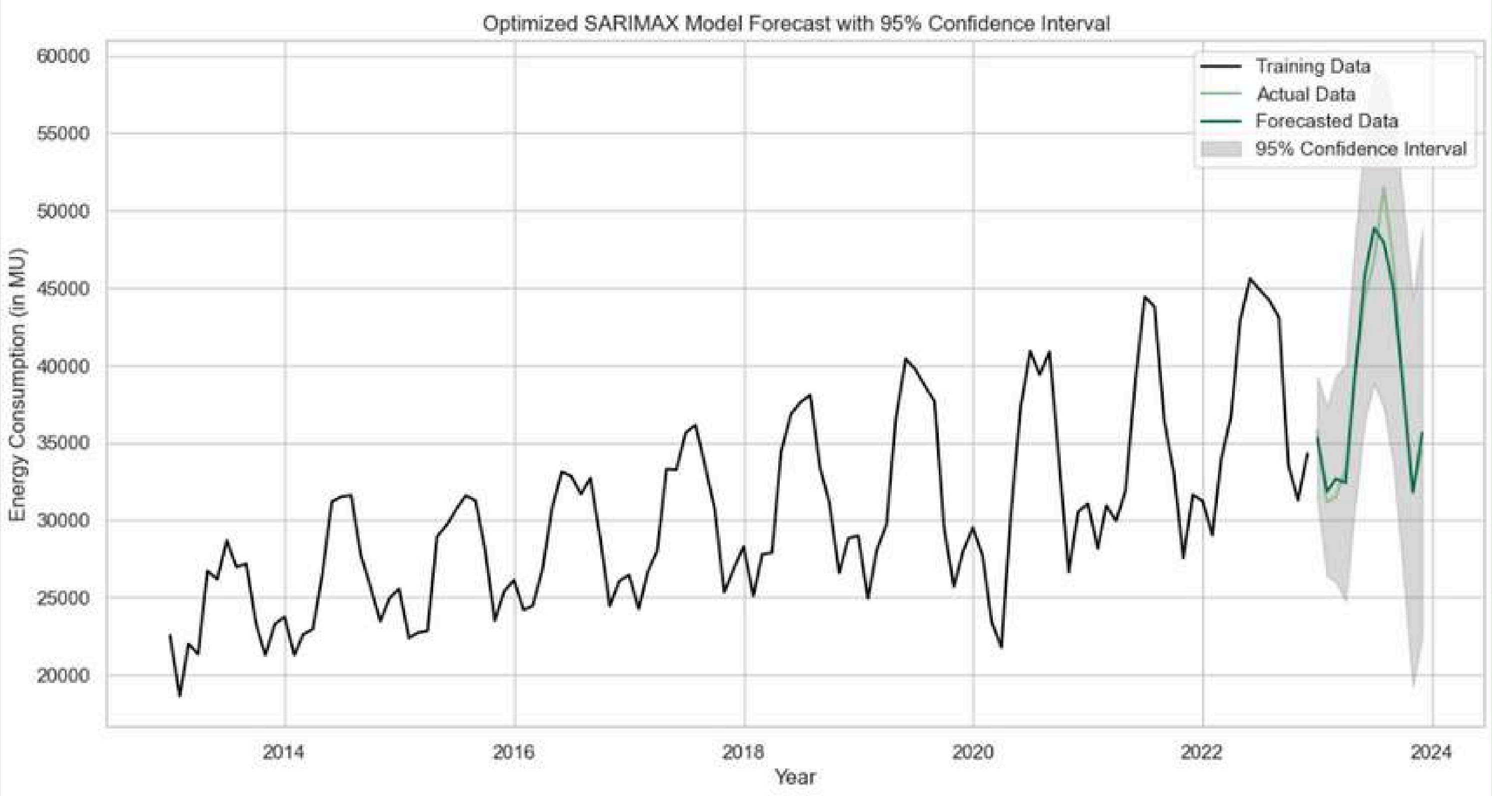
The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.2650Y_{t-2} - 0.5048Y_{t-24} - 1.0601\epsilon_{t-12} + 0.6221\epsilon_{t-24}$$

where:

- Y_{t-2} and Y_{t-24} represent the actual values at time $t-2$ and $t-24$, respectively.
- ϵ_{t-12} and ϵ_{t-24} are the error terms at time $t-12$ and $t-24$, respectively.
- μ is the constant term (if applicable).

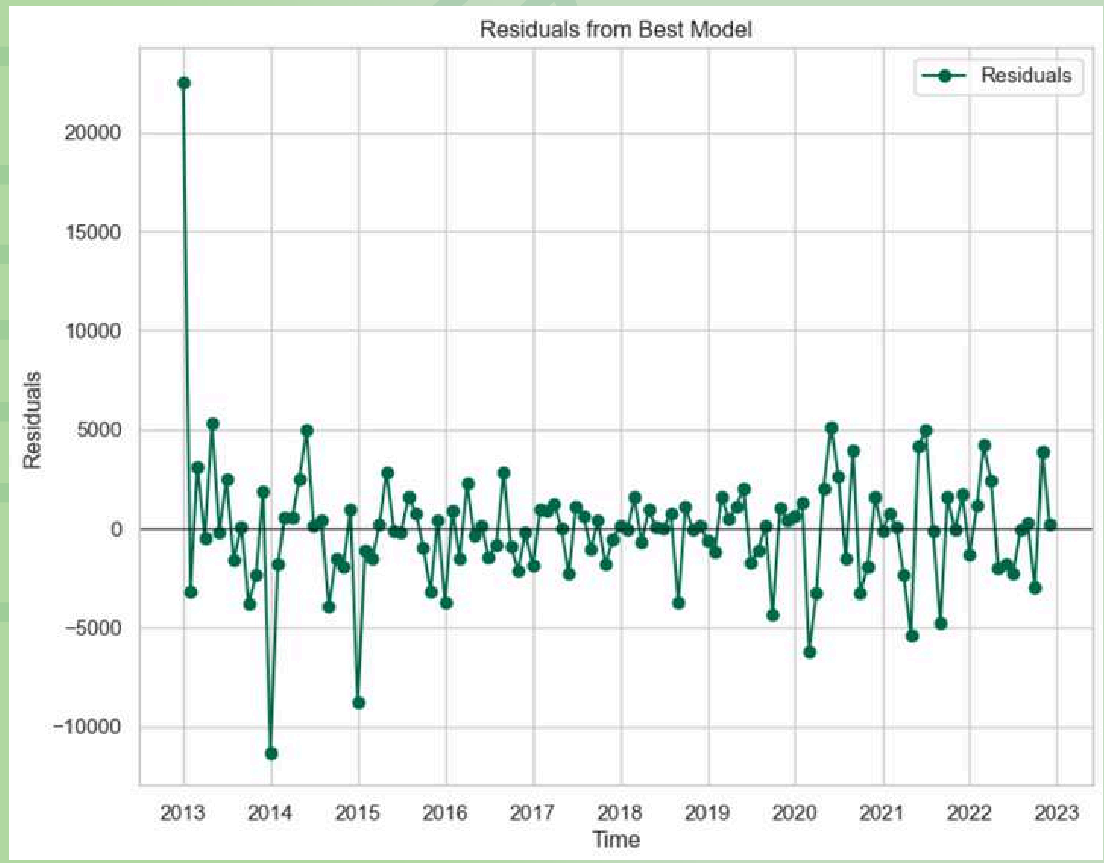
SARIMA Model Fit and Forecasts

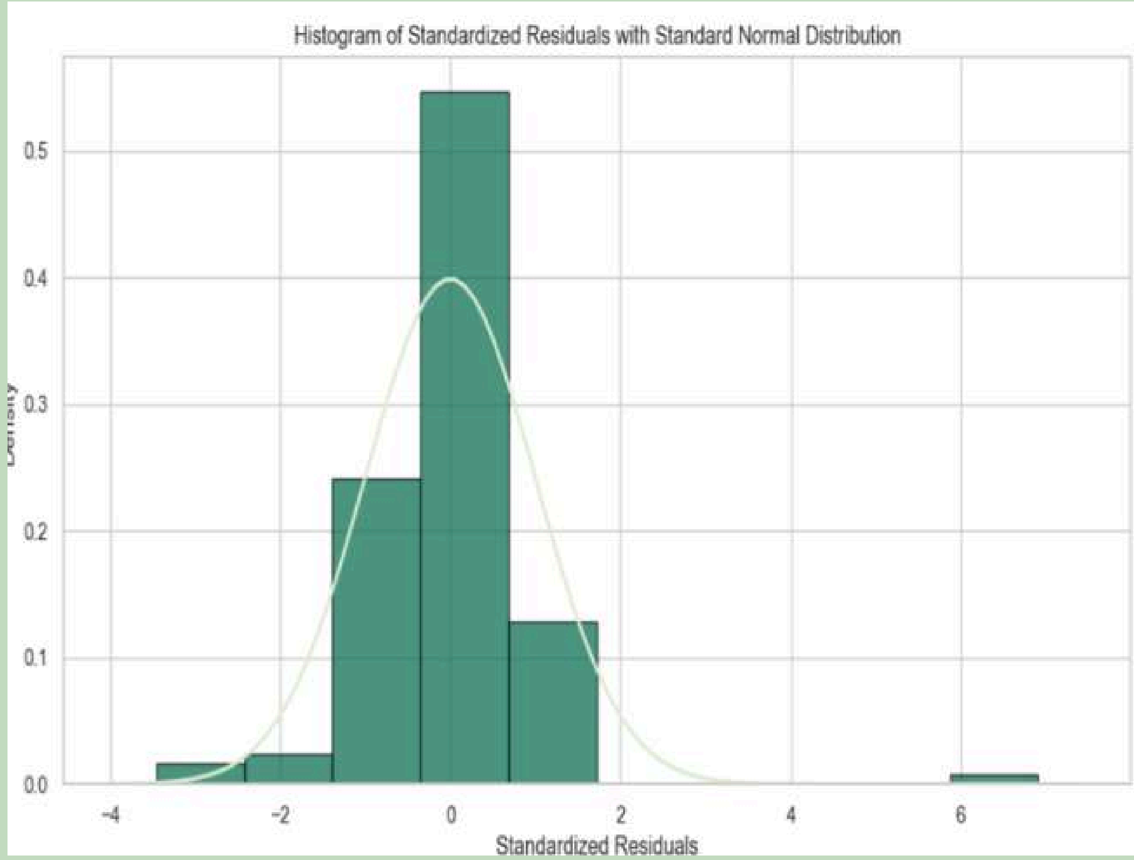


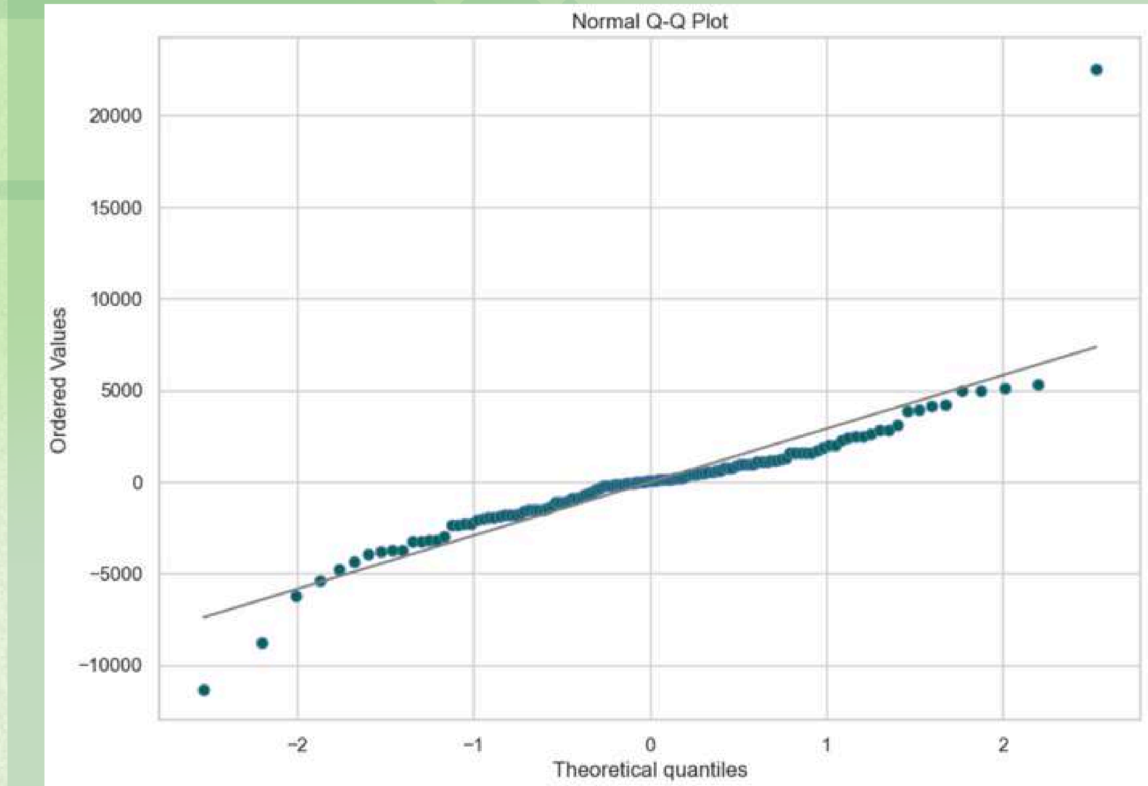
SARIMA Model Accuracy Metrics

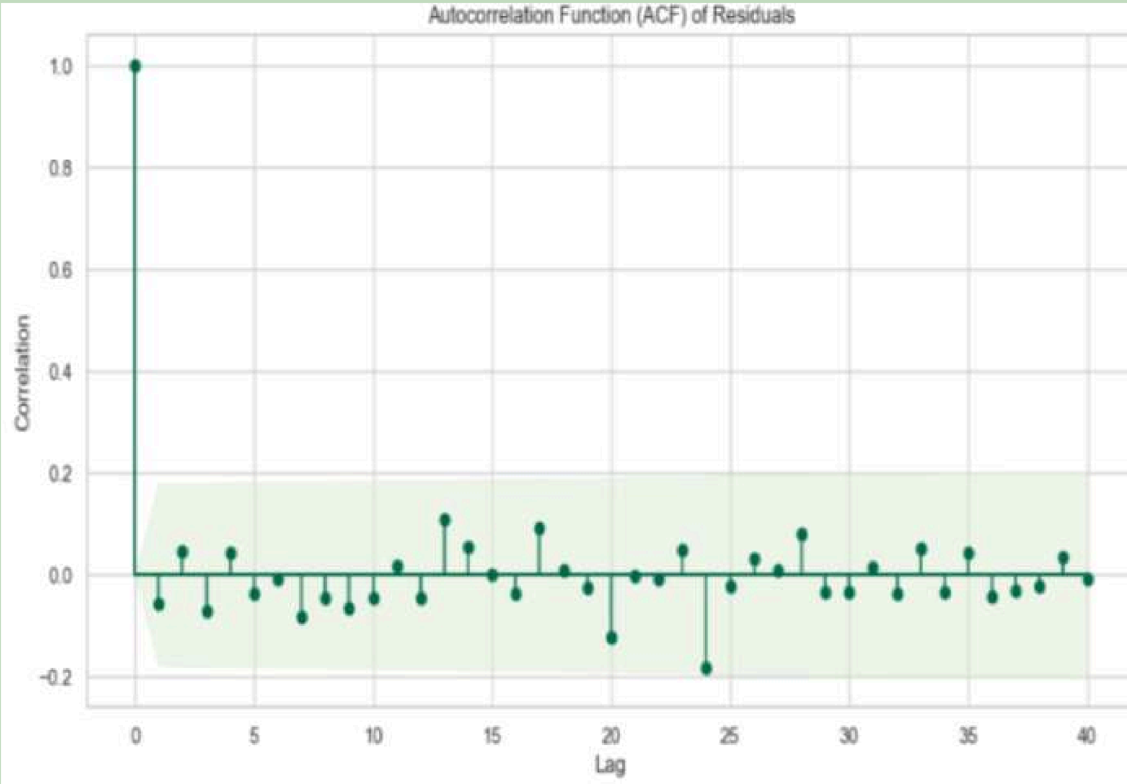
MAPE (Mean Absolute Percentage Error)	**3.46%**
MSE (Mean Squared Error)	**3497525.4000322**

Residual Analysis for North









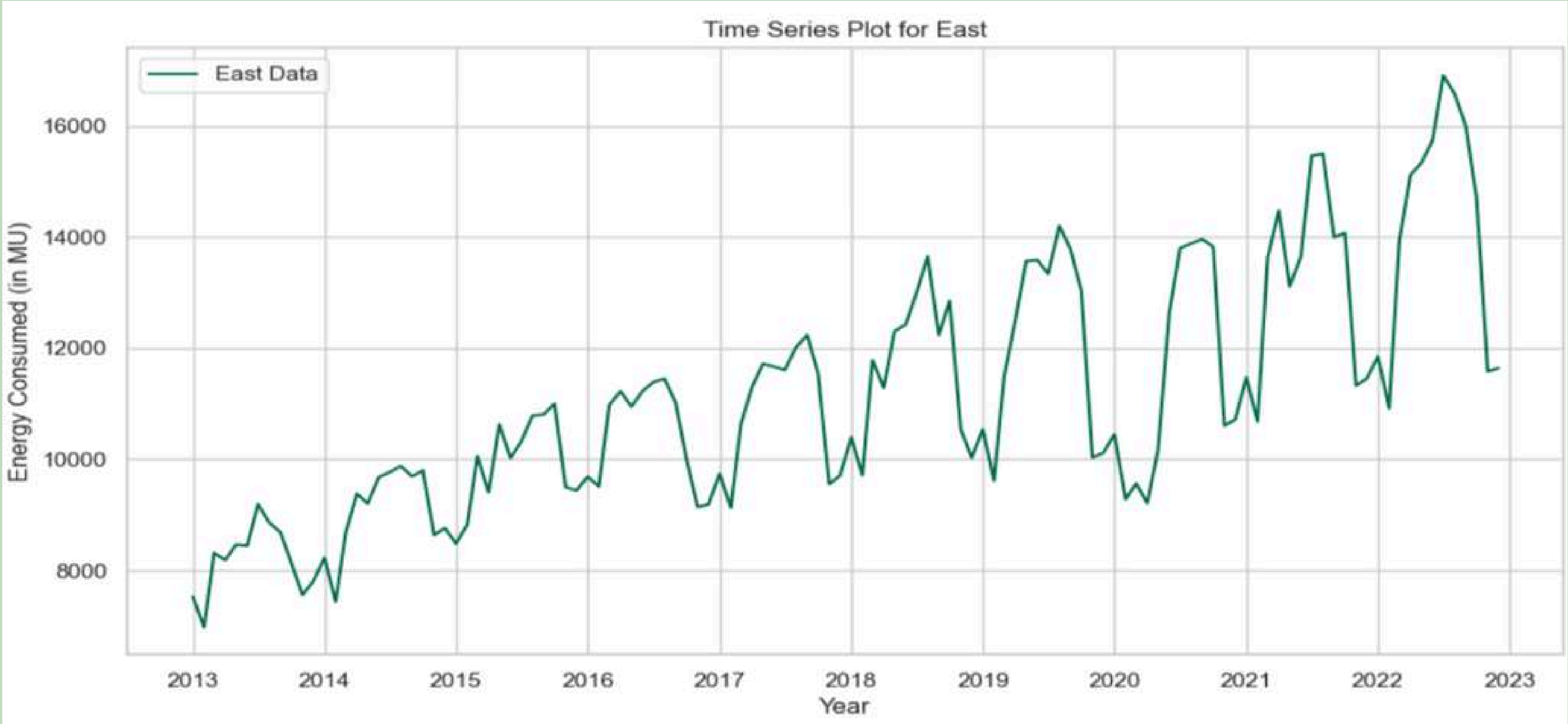
The residual analysis confirms that the SARIMA model satisfies all key assumptions:

- **Normality**: Residuals are normally distributed.
- **Stationarity**: No trends or patterns in the residuals.
- **No Autocorrelation**: Residuals show no significant autocorrelation.
- **Homoscedasticity**: Residuals have constant variance.

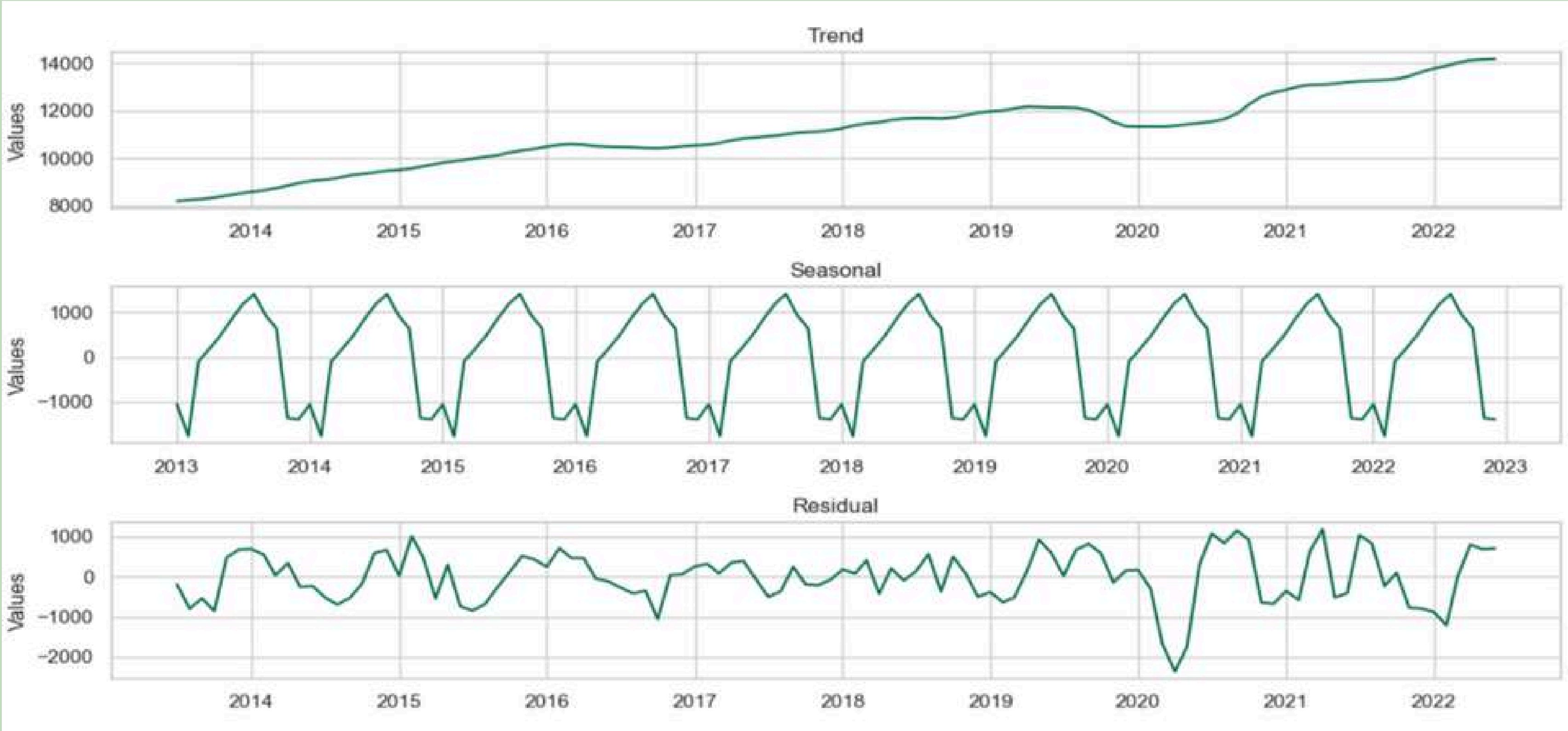
This indicates a good model fit for accurate forecasting.

Electricity Consumption Analysis for East Region

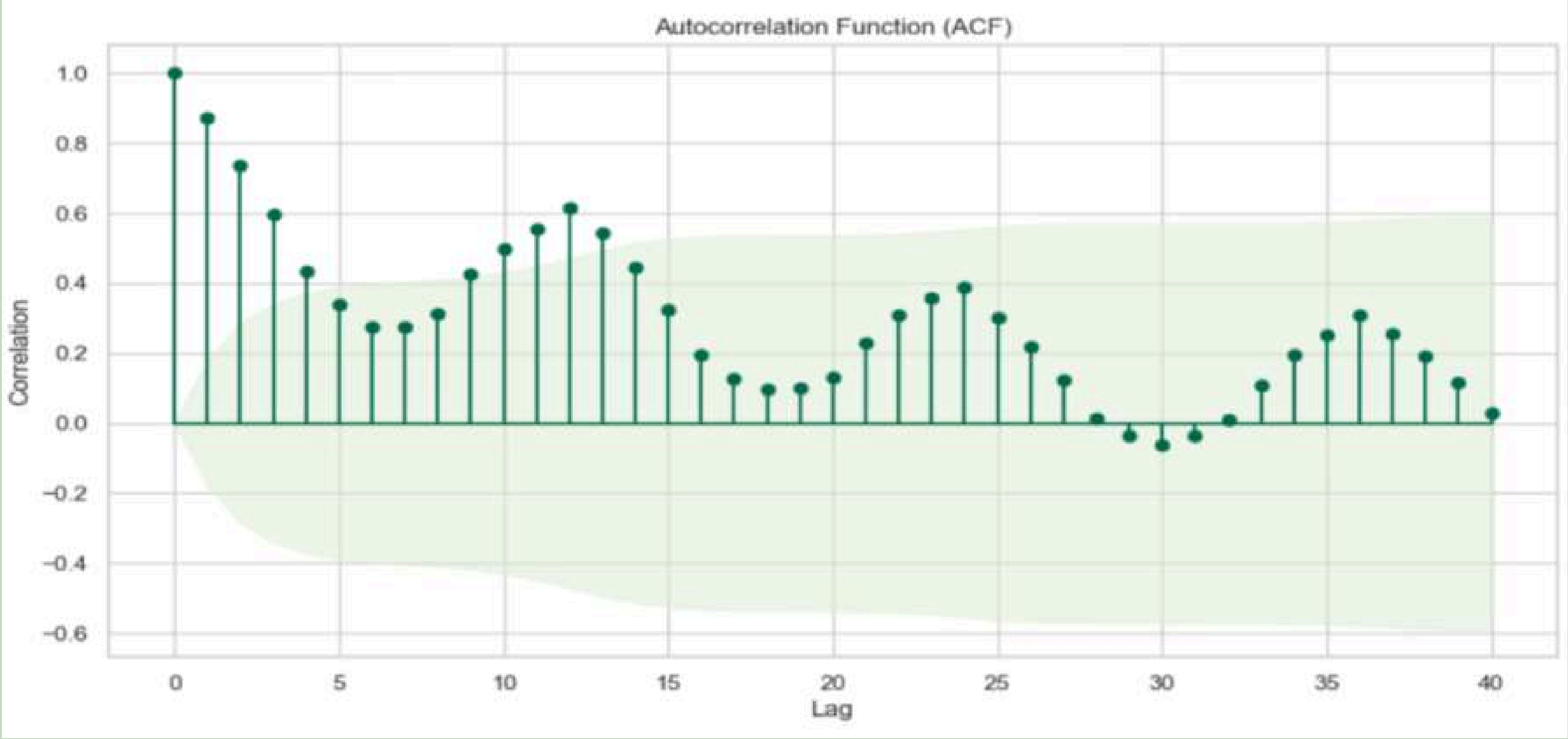
Time Series Plot



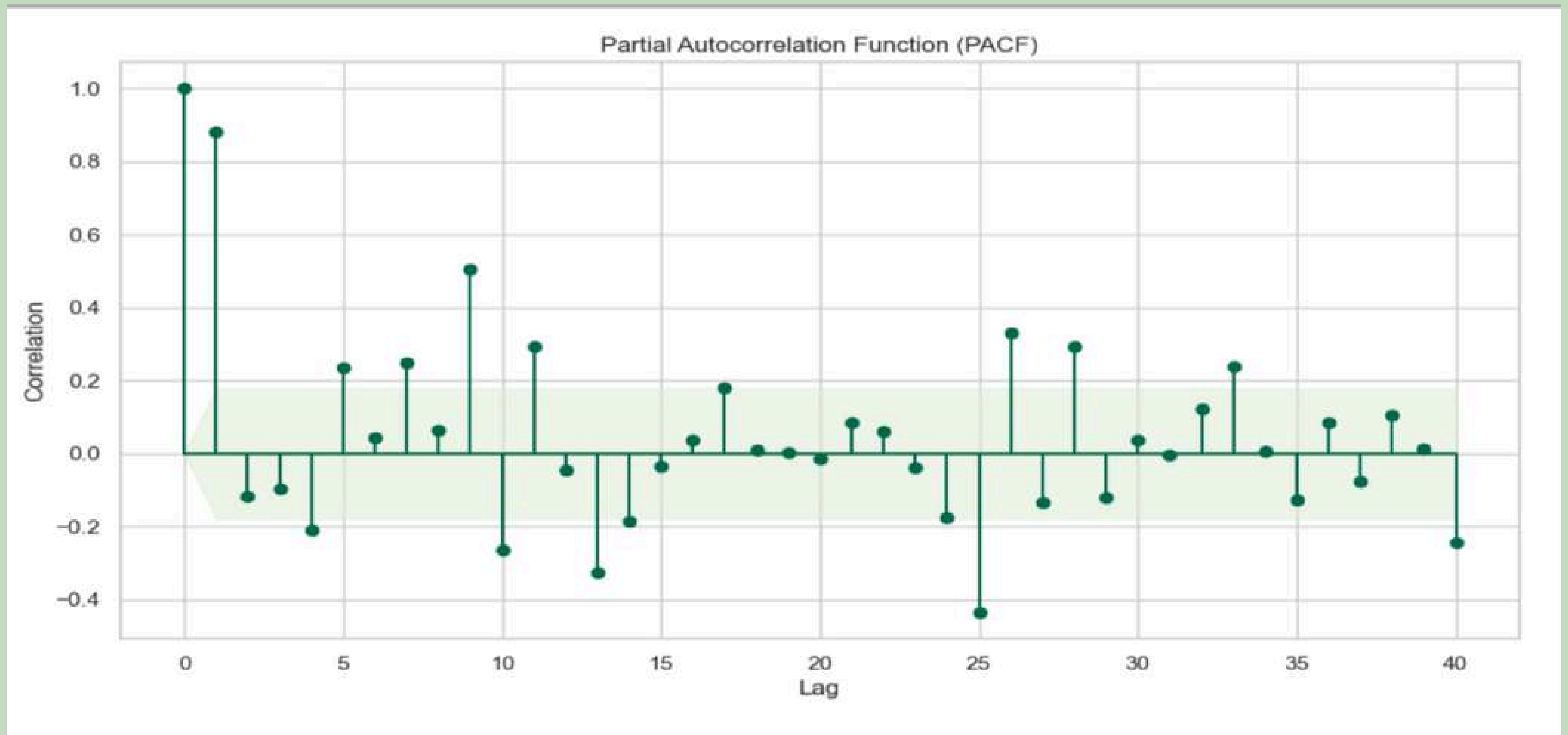
Time Series Decomposition Plot



Autocorrelation Function (ACF) Plot



Partial Autocorrelation Function (PACF) Plot



Stationarity Check

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(train_data)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -0.502700
p-value: 0.891489
```

- Negative ADF Statistic
- A p-value of 0.891489 (which is greater than L.o.S 0.05)

Thus, we fail to reject the Null Hypothesis of the data being non-stationary and can safely conclude that our data is non-stationary.

Also, from the decomposition performed, we can also conclude that seasonality is present in the data at

Preprocessing for Stationarity: Seasonal and Normal Differencing

After Seasonal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(sd_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -2.278178
p-value: 0.179103
```

After Normal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(double_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -6.714803
p-value: 0.000000
```

The time series data is stationary after applying seasonal and normal differencing, as indicated by the p-value being less than 0.05 from the ADF test.

Holt-Winters Linear Exponential Smoothing Model Results

Holt-Winters Model Summary

Dep. Variable:	East	**No. Observations:**	120
Model:	ExponentialSmoothing	**SSE**	77909903.746
Optimized:	True	**AIC**	1638.029
Trend:	Additive	**BIC**	1682.628
Seasonal:	Additive	**AICC**	1644.801
Seasonal Periods:	12	**Date:**	Wed, 04 Sep 2024
Box-Cox:	False	**Time:**	04:09:42
Box-Cox Coeff.:	None		

	coeff	code
smoothing_level	7e-05	alpha
smoothing_trend	0.9879675	beta
smoothing_seasonal	0.0006345	gamma

Holt-Winters Model Equations

Holt-Winters Model

The Holt-Winters model equations based on the provided model summary are as follows:

Level Equation

$$L_t = 0.00007(Y_t - S_t) + (1 - 0.00001)(L_{t-1} + T_{t-1})$$

Trend Equation

$$T_t = 0.9879675(L_t - L_{t-1}) + (1 - 0.9879675)T_{t-1}$$

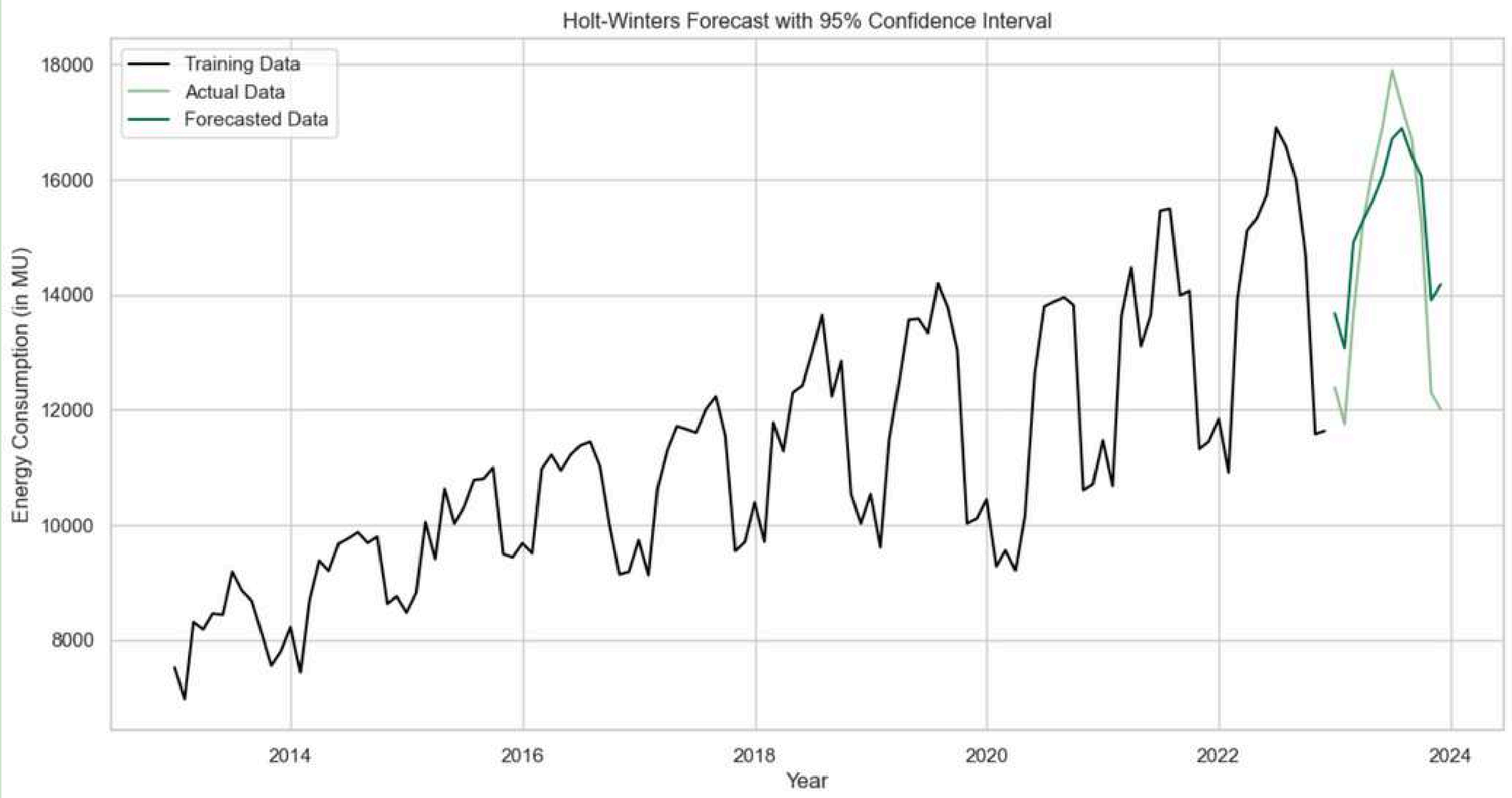
Seasonal Component

$$S_t = 0.0006345(Y_t - L_t) + (1 - 0.0006345)S_{t-12}$$

Forecast Equation

$$\hat{Y}_{t+h} = L_t + h \cdot T_t + S_{t+12-h}$$

Holt-Winters Model Fit and Forecasts



Holt-Winters Model Accuracy Metrics

MAPE (Mean Absolute Percentage Error) → **6.95%**

MSE (Mean Squared Error) → **1436043.5774238**

SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Summary

Dep. Variable:	East	No. Observations:	120
Model:	SARIMAX(2, 1, 2)x(0, 1, 0, 12)	Log Likelihood	-839.803
Date:	Wed, 04 Sep 2024	AIC	1689.605
Time:	09:35:44	BIC	1702.827
Sample:	01-01-2013	HQIC	1694.962
	- 12-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.2132	0.070	-3.064	0.002	-0.350	-0.077
ar.L2	0.6234	0.071	8.832	0.000	0.485	0.762
ma.L1	-0.0004	2.952	-0.000	1.000	-5.786	5.785
ma.L2	-0.9996	0.122	-8.160	0.000	-1.240	-0.759
sigma2	5.596e+05	5.25e-06	1.07e+11	0.000	5.6e+05	5.6e+05

Ljung-Box (L1) (Q):	0.16	Jarque-Bera (JB):	4.45
Prob(Q):	0.69	Prob(JB):	0.13
Heteroskedasticity (H):	2.55	Skew:	0.42
Prob(H) (two-sided):	0.01	Kurtosis:	4.39

SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\mathrm{SARIMA}(2,1,2)(0,1,0)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.2132Y_{t-1} + 0.6234Y_{t-2} + 0.9996\epsilon_{t-2}$$

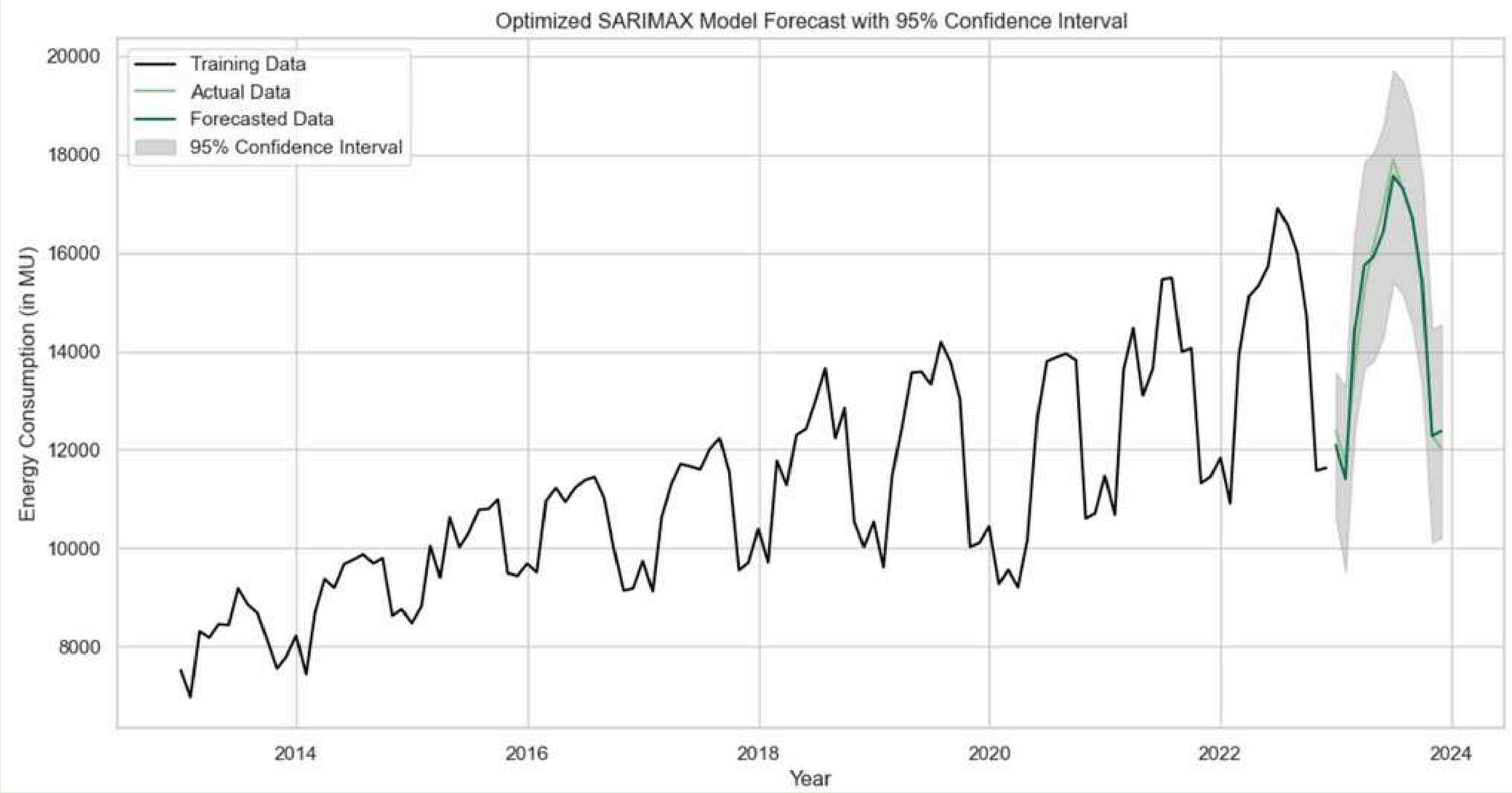
where:

- Y_{t-1} and Y_{t-2} represent the actual values at time $t-1$ and $t-2$, respectively.
- ϵ_{t-2} is the error term at time $t-2$.
- μ is the constant term (if applicable).

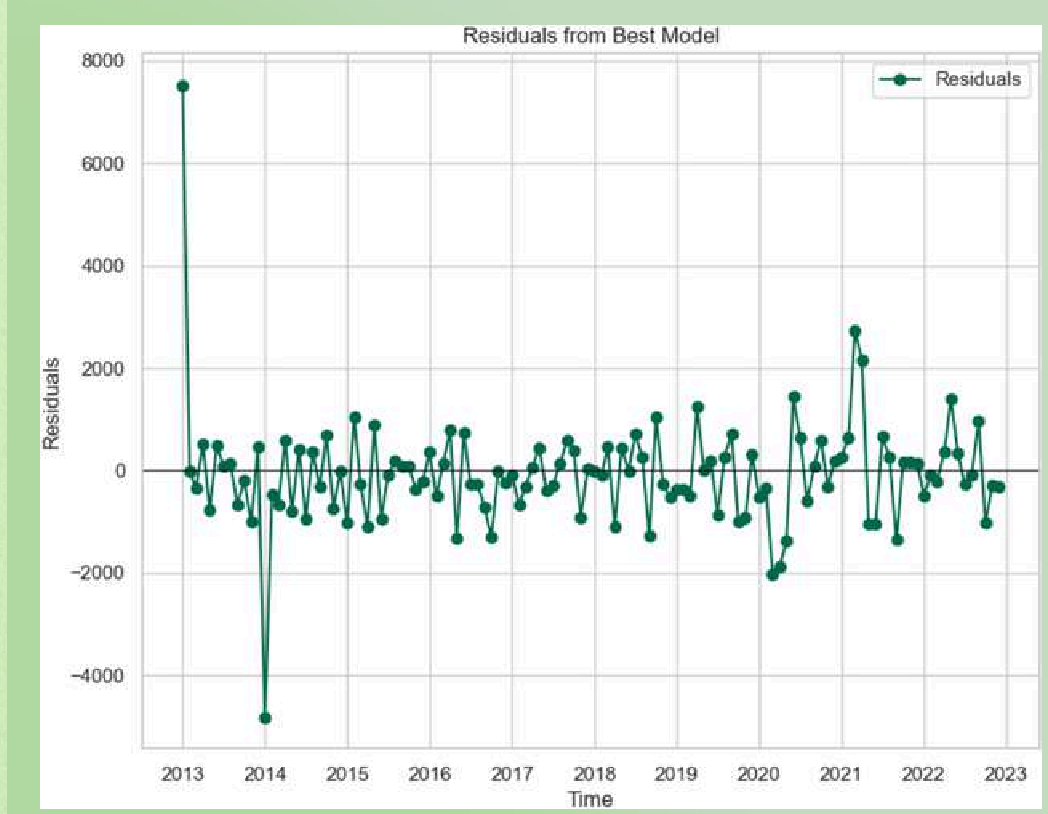
SARIMA Model Accuracy Metrics

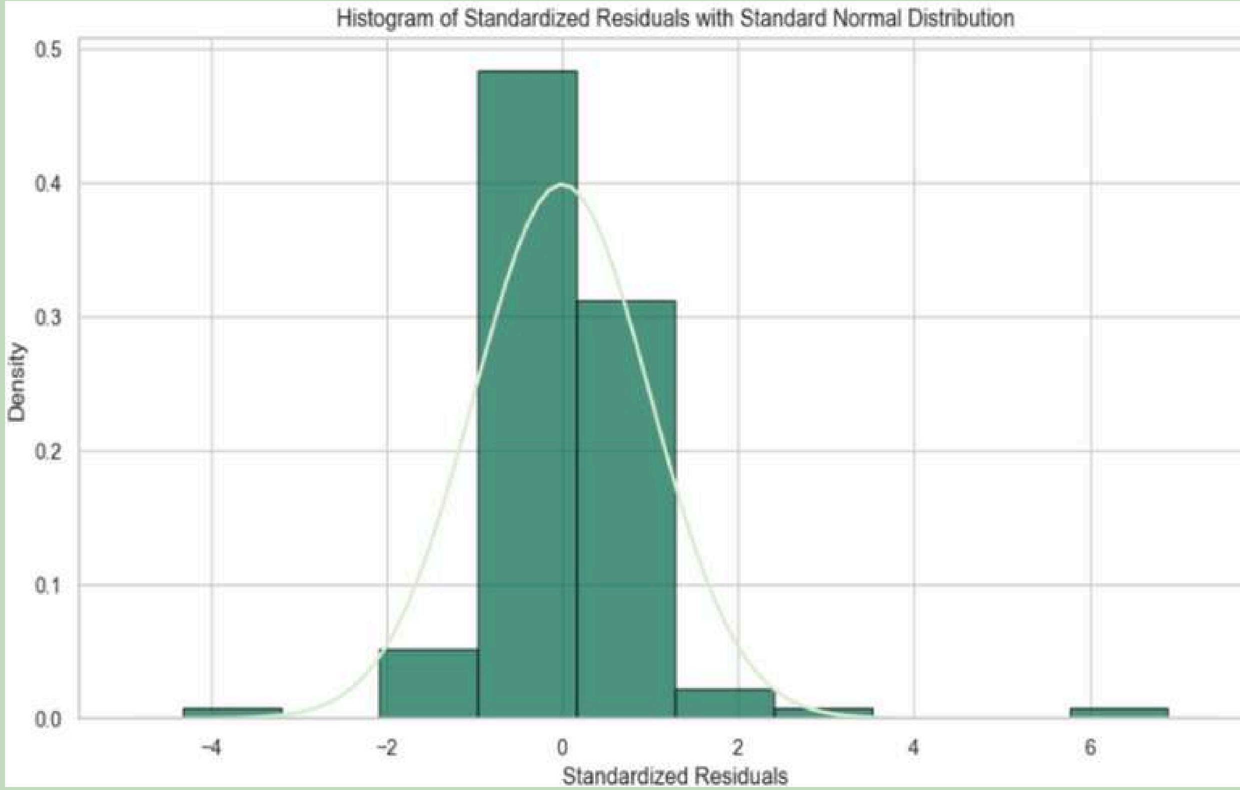
MAPE (Mean Absolute Percentage Error)	**2.10%**
MSE (Mean Squared Error)	**134188.3166541**

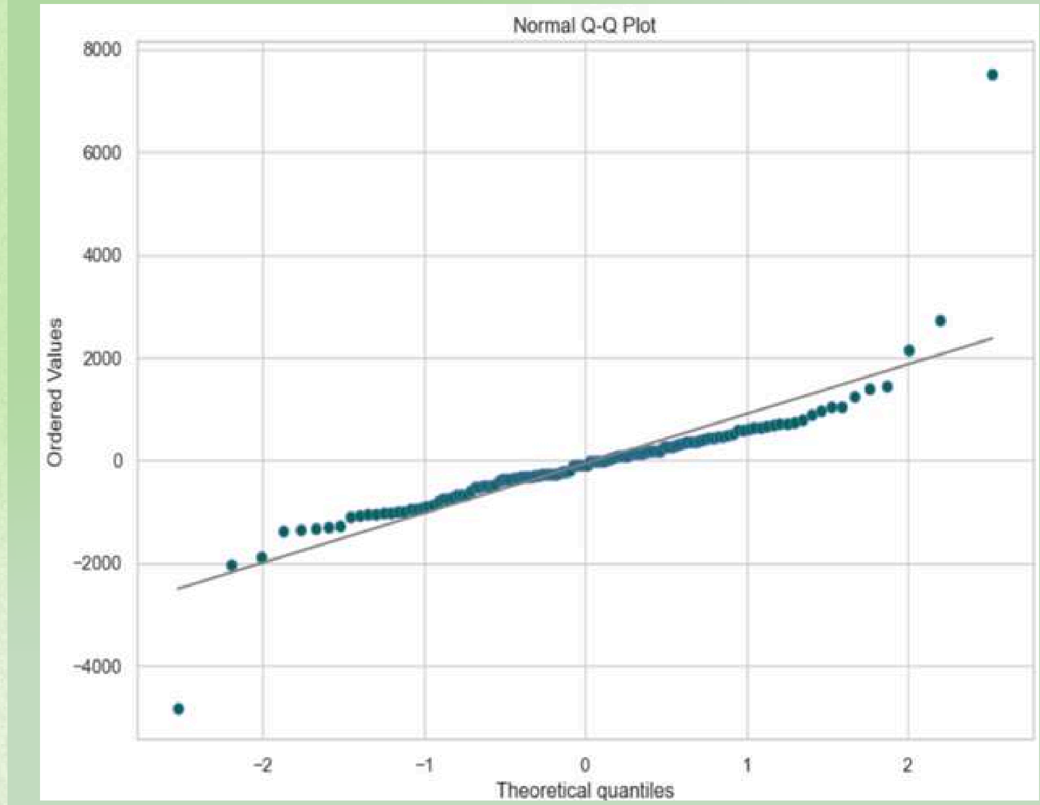
SARIMA Model Fit and Forecasts

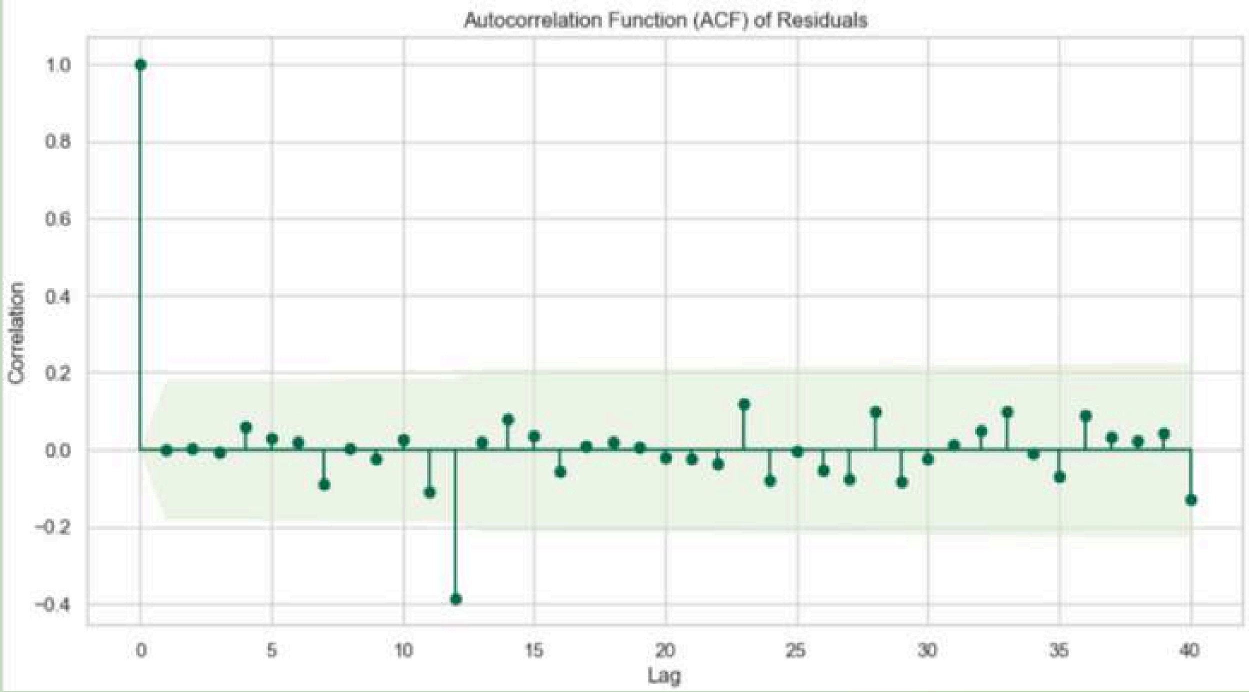


Residual Analysis









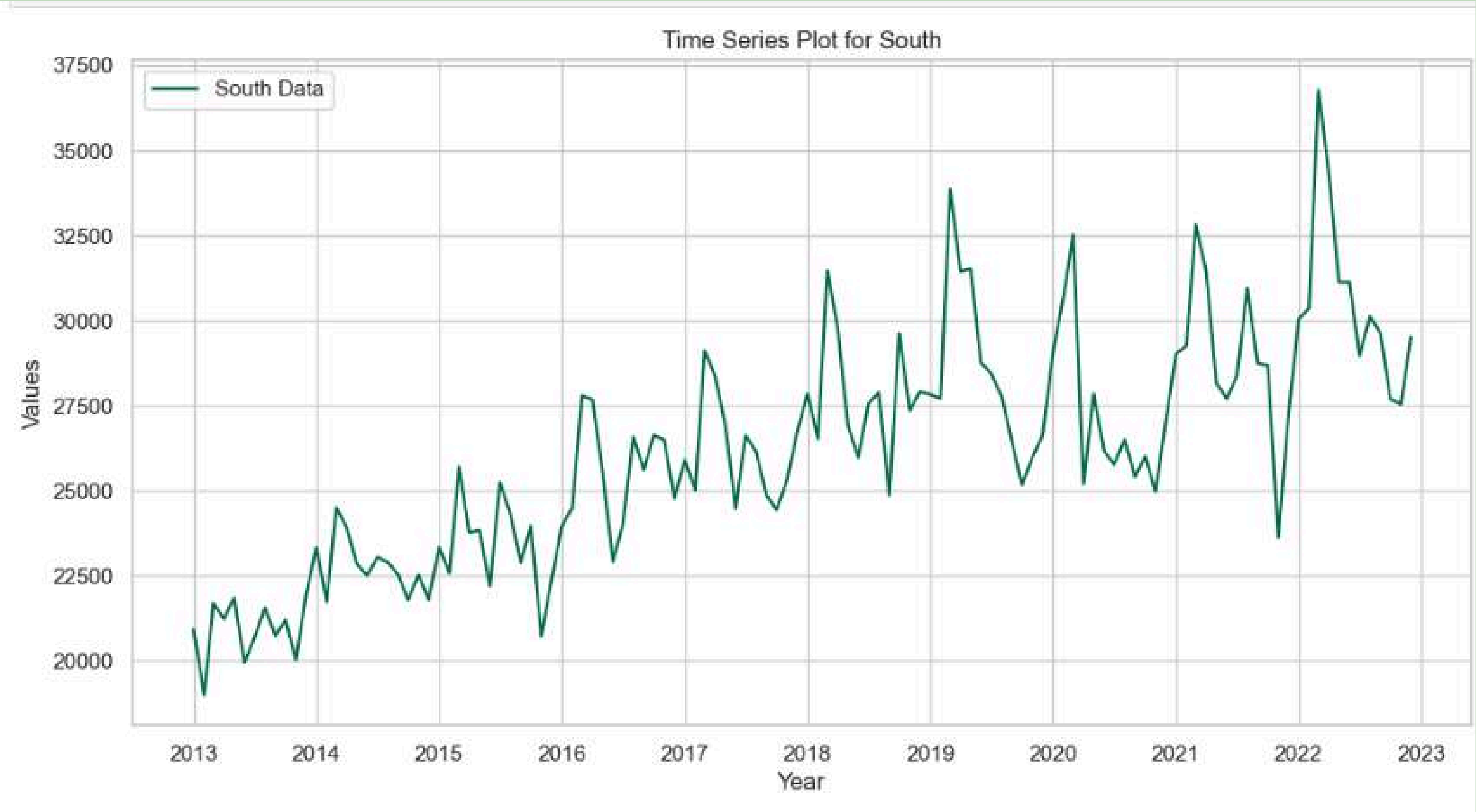
The residual analysis confirms that the SARIMA model satisfies all key assumptions:

- **Normality**: Residuals are normally distributed.
- **Stationarity**: No trends or patterns in the residuals.
- **No Autocorrelation**: Residuals show no significant autocorrelation.
- **Homoscedasticity**: Residuals have constant variance.

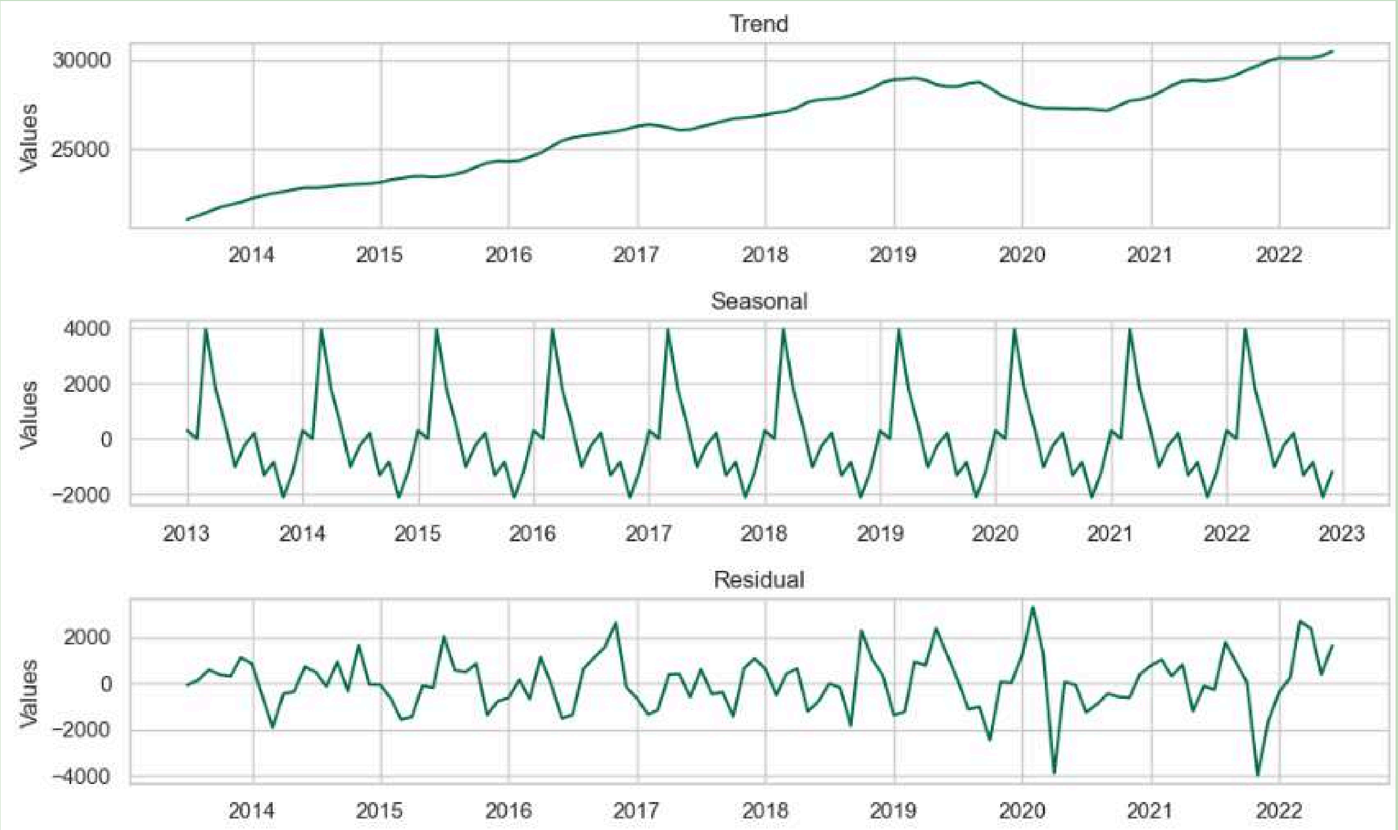
This indicates a good model fit for accurate forecasting.

Electricity Consumption Analysis for South Region

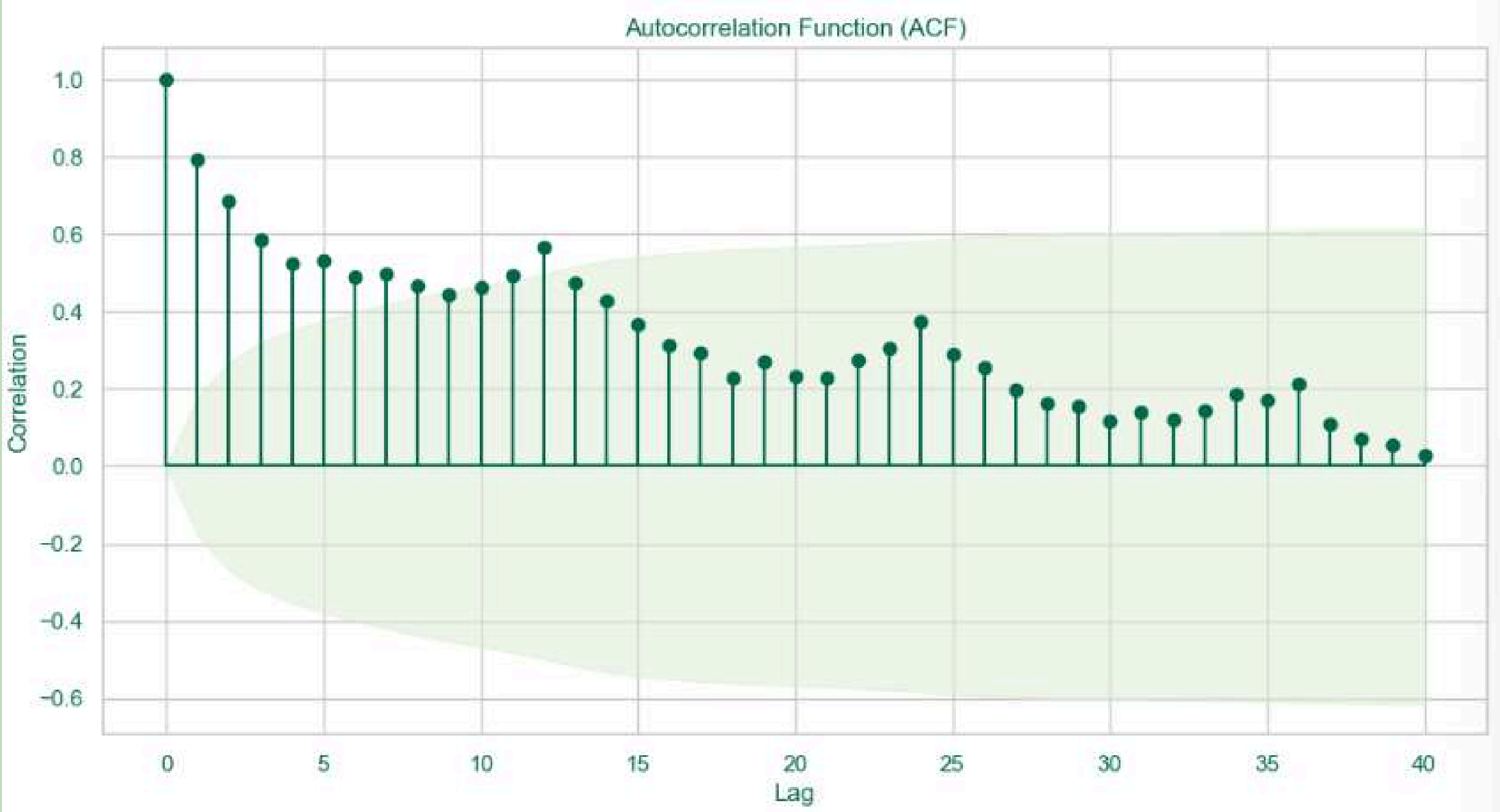
Time Series Plot



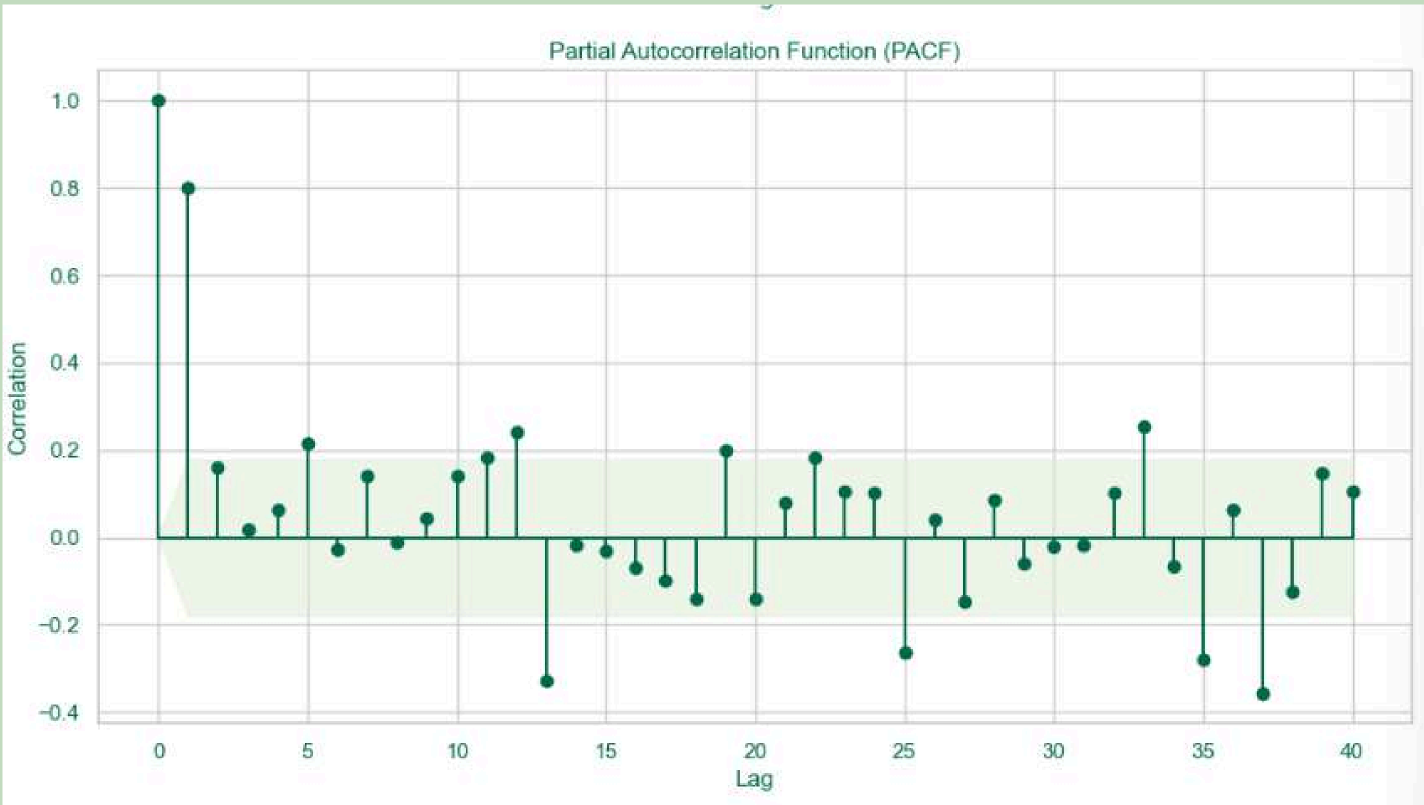
Time Series Decomposition Plot



Autocorrelation Function (ACF) Plot



Partial Autocorrelation Function (PACF) Plot



Stationarity Check

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(train_data)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -1.052568
p-value: 0.733612
```

- Negative ADF Statistic
- A p-value of 0.733612 (which is greater than L.o.S 0.05)

Thus, we fail to reject the null hypothesis that the data is non-stationary and can safely conclude that our data is non-stationary.

Also, from the decomposition performed, we can also conclude that seasonality is present in the data

Preprocessing for Stationarity: Seasonal and Normal Differencing

After Seasonal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
sd_df_train = train_data - train_data.shift(12)
sd_df_train = sd_df_train.dropna()
result = adfuller(sd_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -3.089693
p-value: 0.027325
```

After Normal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
double_df_train= sd_df_train - sd_df_train.shift(1)
double_df_train= double_df_train.dropna()
result = adfuller(double_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -3.064903
p-value: 0.029262
```

In both instances, differencing indicates stationarity with a p-value of less than 0.05, confirming that the data is stationary.

Holt-Winters Linear Exponential Smoothing Model Results

Holt-Winters Model Summary

Dep. Variable:	South	**No. Observations:**	120
Model:	ExponentialSmoothing	**SSE**	249201663.128
Optimized:	True	**AIC**	1777.554
Trend:	Additive	**BIC**	1822.154
Seasonal:	Additive	**AICC**	1784.326
Seasonal Periods:	12	**Date:**	Wed, 04 Sep 2024
Box-Cox:	False	**Time:**	01:13:14
Box-Cox Coeff.:	None		

	coeff	code	optimized
smoothing_level	0.6060714	alpha	True
smoothing_trend	0.0001	beta	True
smoothing_seasonal	0.0001	gamma	True

Holt-Winters Model Equations

Holt-Winters Model

The Holt-Winters model equations based on the provided model summary are as follows:

Level Equation

$$L_t = 0.6060714(Y_t - S_t) + (1 - 0.6060714)(L_{t-1} + T_{t-1})$$

Trend Equation

$$T_t = 0.0001(L_t - L_{t-1}) + (1 - 0.0001)T_{t-1}$$

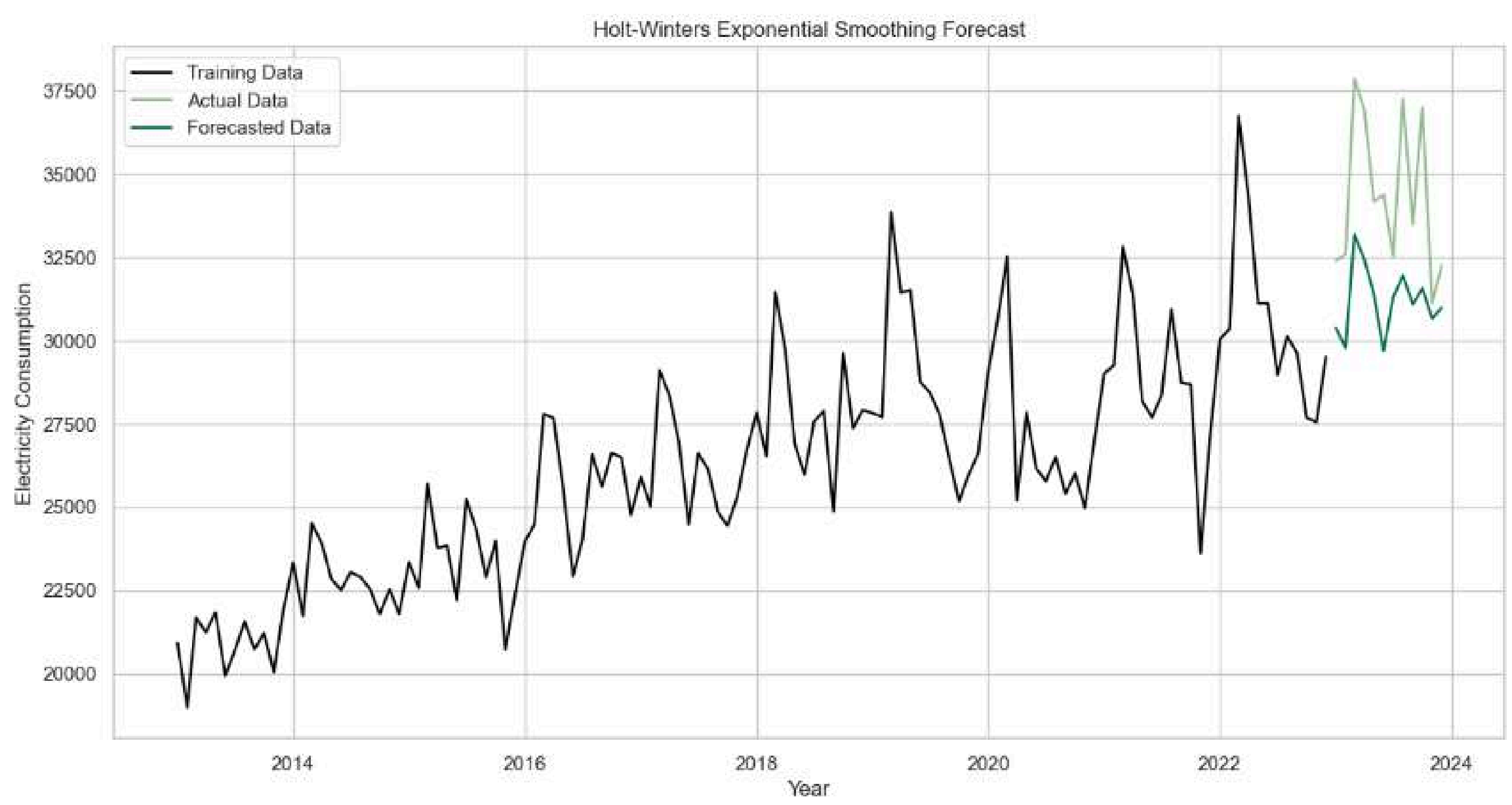
Seasonal Component

$$S_t = 0.0001(Y_t - L_t) + (1 - 0.0001)S_{t-12}$$

Forecast Equation

$$\hat{Y}_{t+h} = L_t + h \cdot T_t + S_{t+12-h}$$

Holt-Winters Model Fit and Forecasts



Holt-Winters Model Accuracy Metrics

MAPE (Mean Absolute Percentage Error) → **8.88%**

MSE (Mean Squared Error) → **12565414.52859798**

SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Summary

SARIMAX Results

Dep. Variable:	South	**No. Observations:**	120
Model:	SARIMAX(1, 1, 0)x(0, 1, 0, 12)	**Log Likelihood**	-955.202
Date:	Wed, 04 Sep 2024	**AIC**	1914.403
Time:	01:33:30	**BIC**	1919.730
Sample:	01-01-2013	**HQIC**	1916.562
	- 12-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.3501	0.083	-4.239	0.000	-0.512	-0.188
sigma2	3.933e+06	4.31e+05	9.123	0.000	3.09e+06	4.78e+06

Ljung-Box (L1) (Q):	0.22	**Jarque-Bera (JB):**	7.50
Prob(Q):	0.64	**Prob(JB):**	0.02
Heteroskedasticity (H):	2.51	**Skew:**	-0.14
Prob(H) (two-sided):	0.01	**Kurtosis:**	4.27

SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\text{SARIMA}(1,1,0)(0,1,0)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.3501 Y_t - \epsilon_t$$

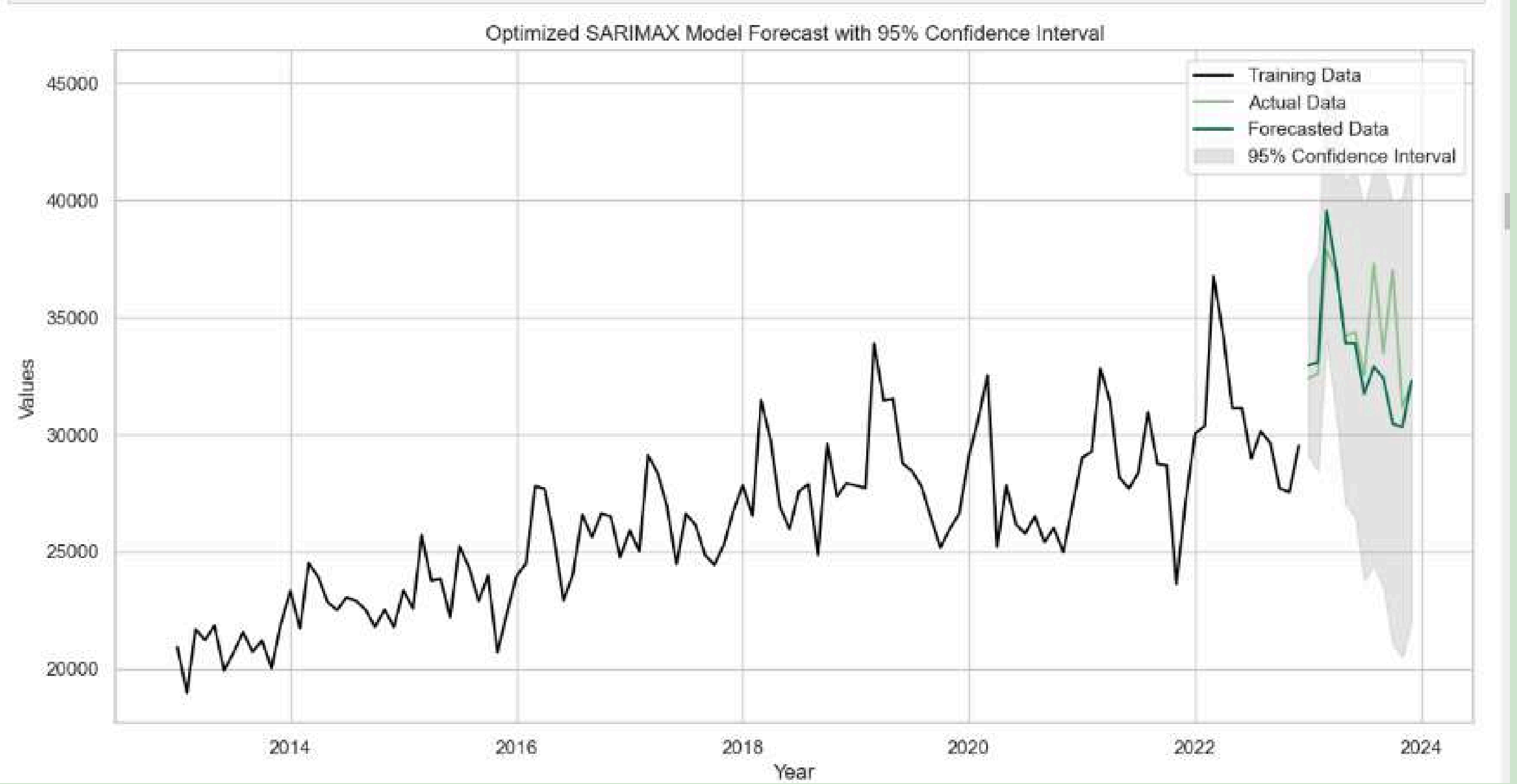
where:

- Y_t represents the actual value at time t.
- ϵ_t is the error term at time t.
- μ is the constant term (if applicable).

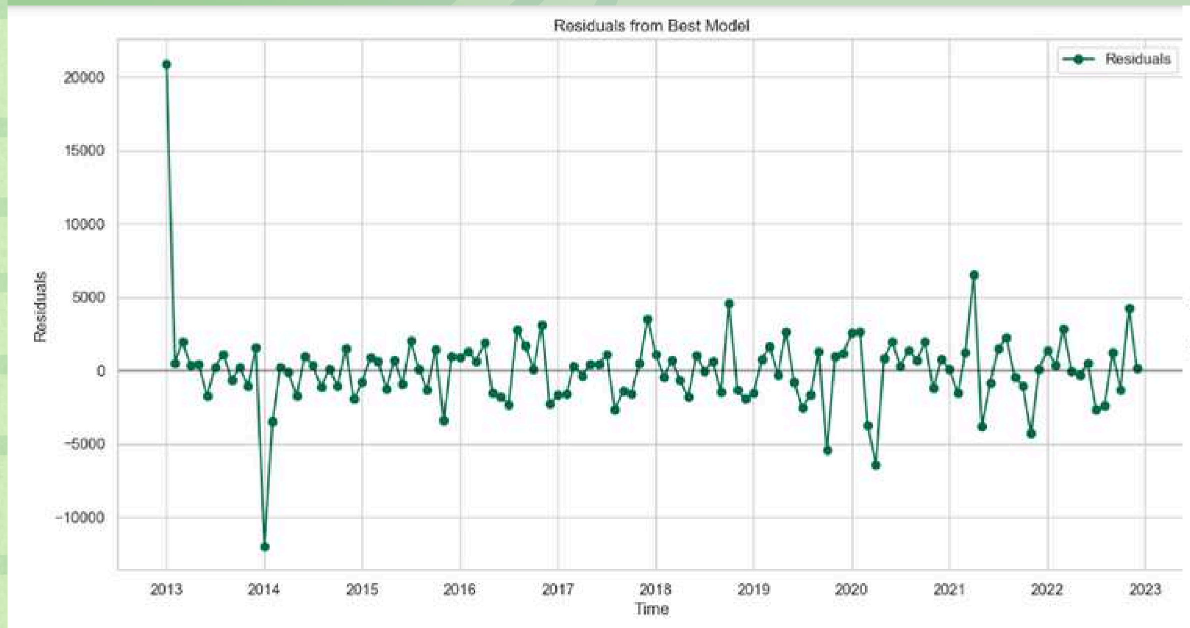
SARIMA Model Accuracy Metrics

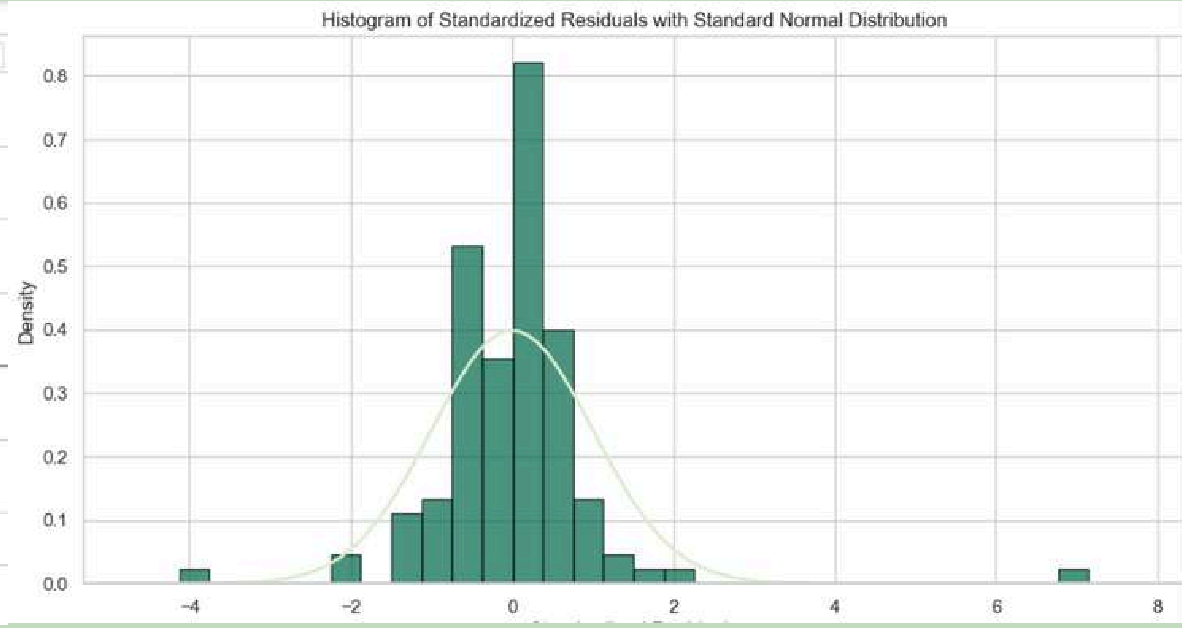
MAPE (Mean Absolute Percentage Error)	**4.01%**
MSE (Mean Squared Error)	**5696498.1786016**

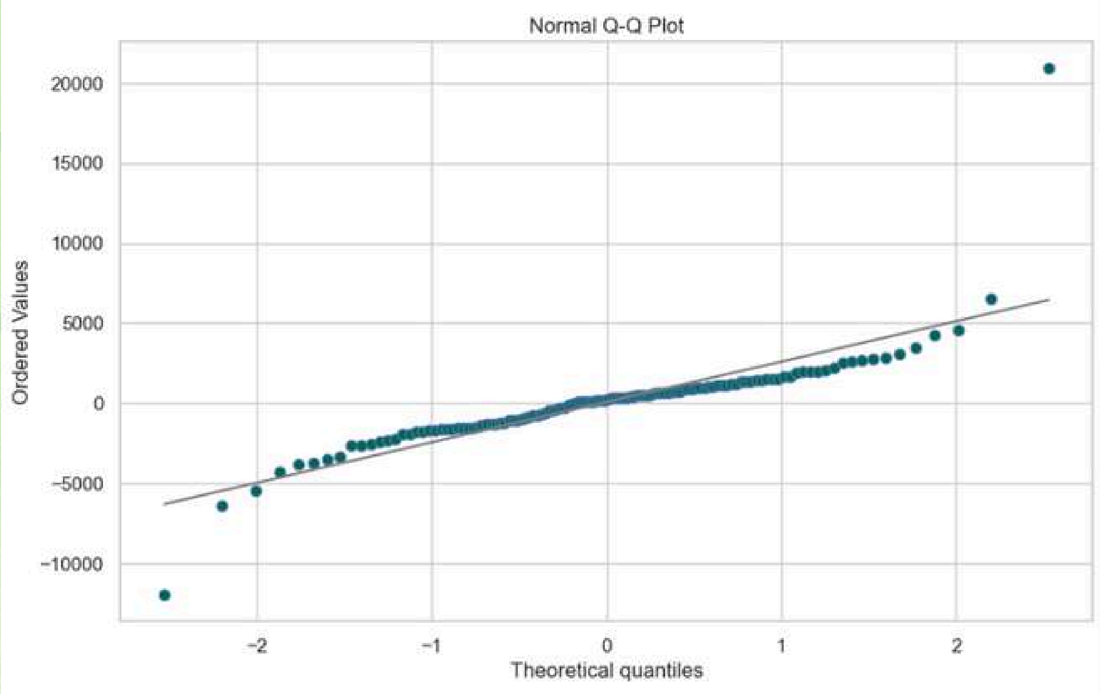
SARIMA Model Fit and Forecasts

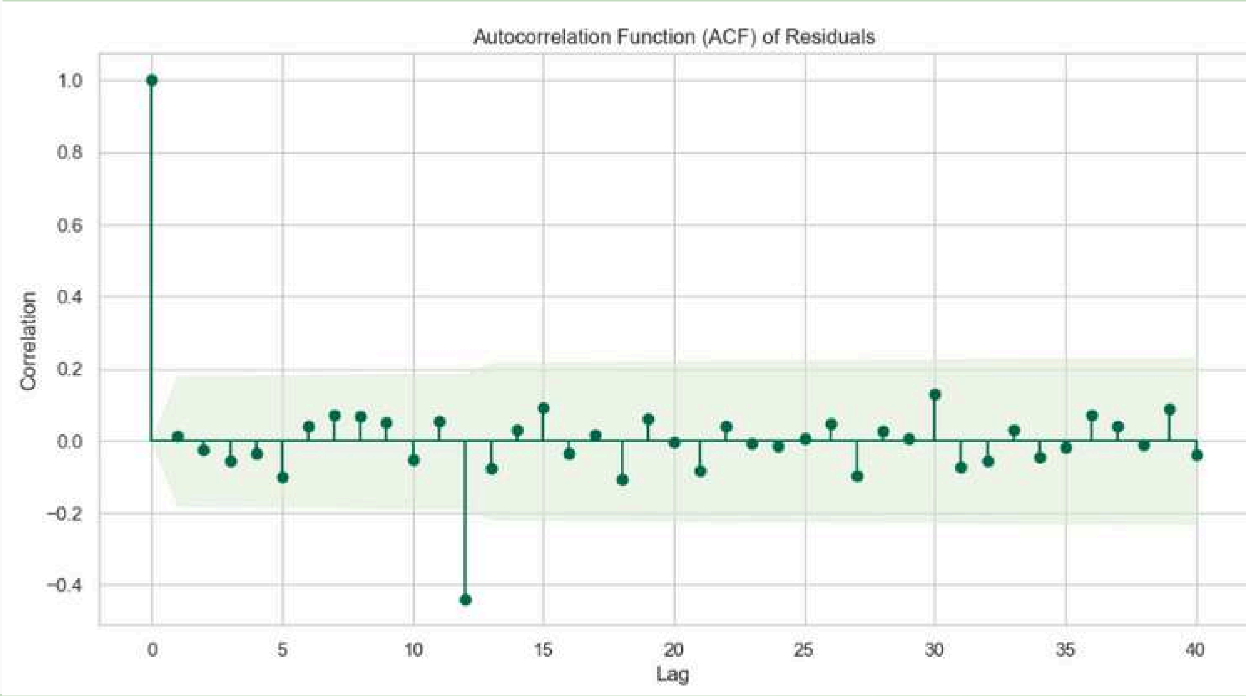


Residual Analysis









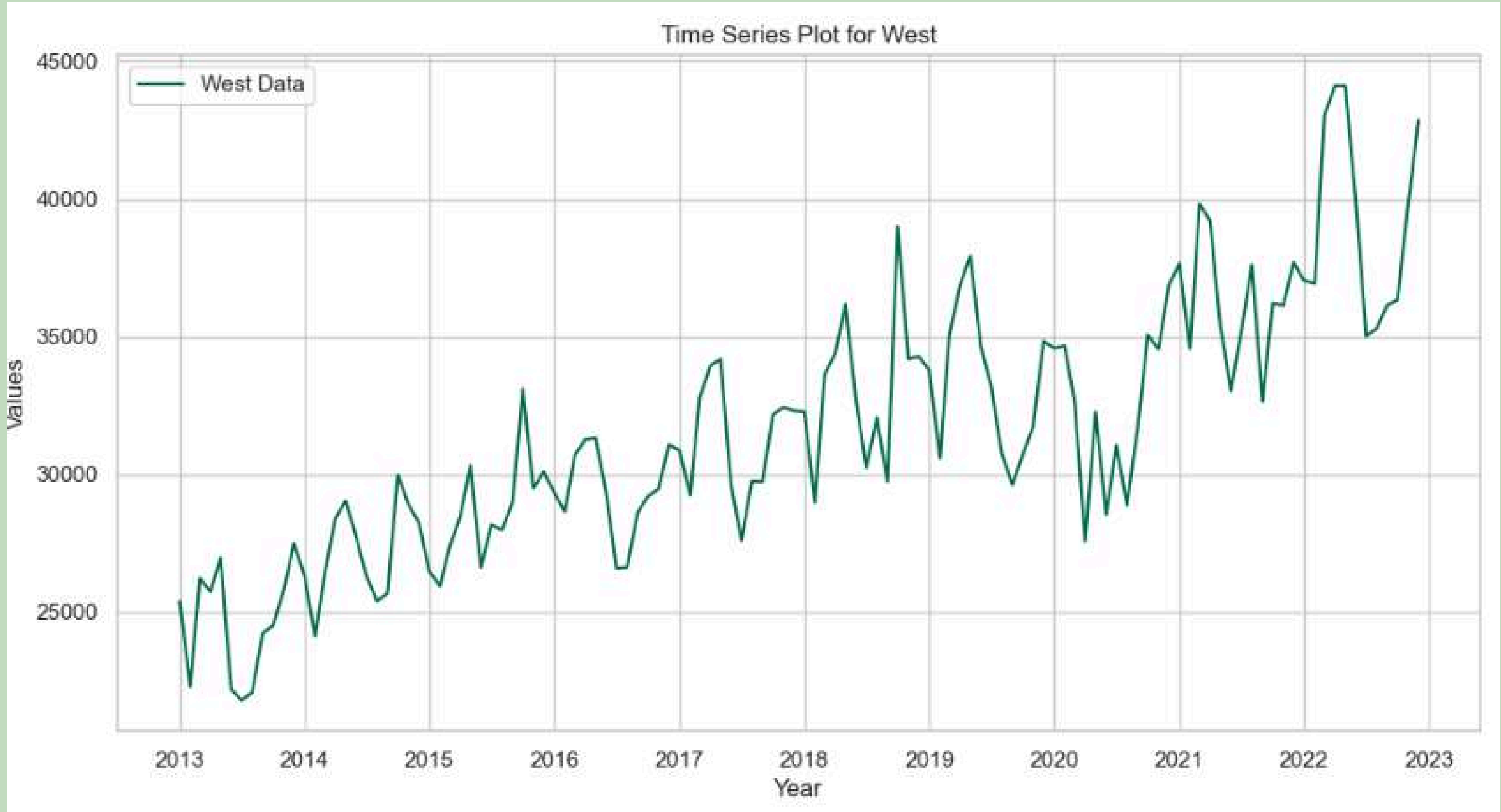
The residual analysis confirms that the SARIMA model satisfies all key assumptions:

- **Normality**: Residuals are normally distributed.
- **Stationarity**: No trends or patterns in the residuals.
- **No Autocorrelation**: Residuals show no significant autocorrelation.
- **Homoscedasticity**: Residuals have constant variance.

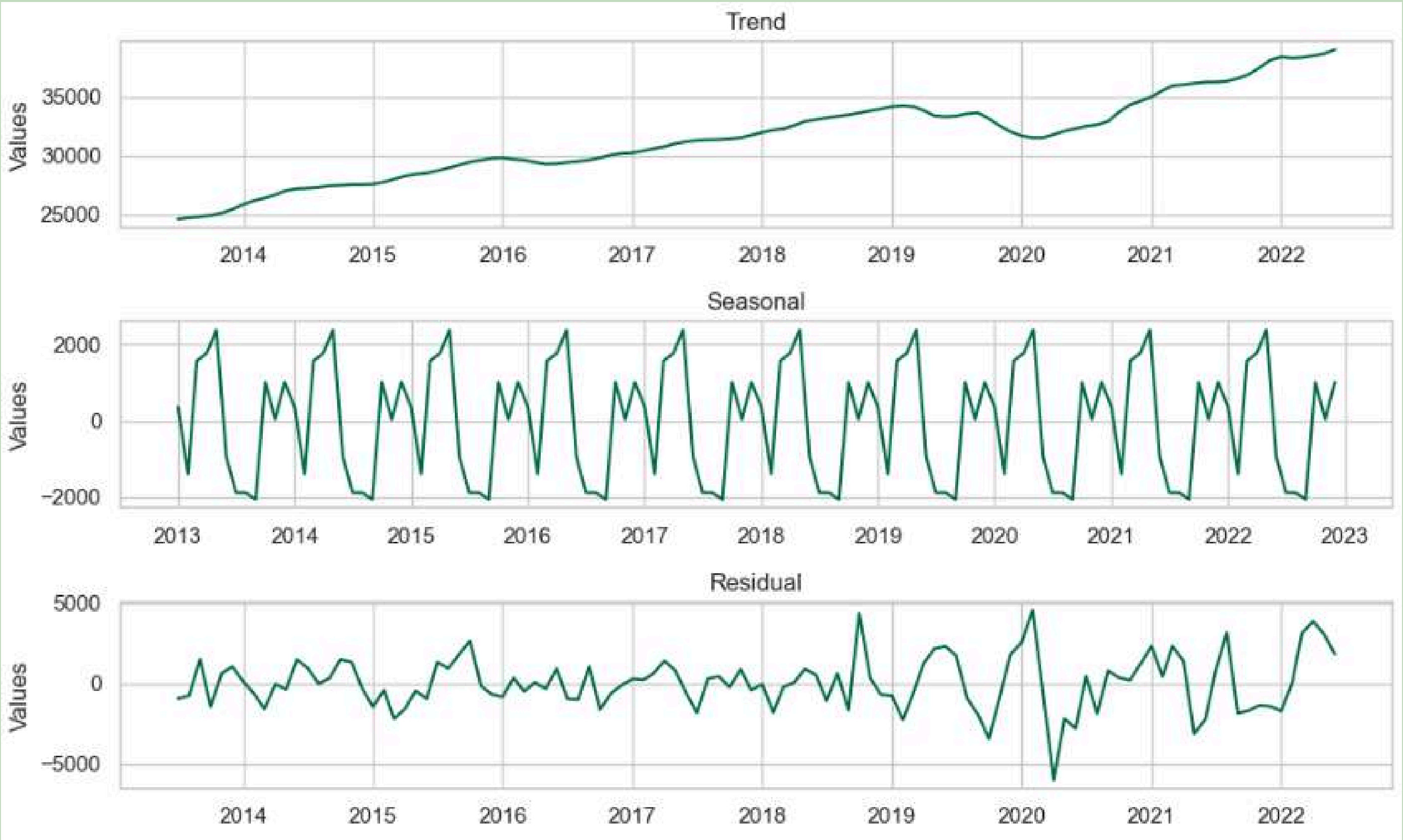
This indicates a good model fit for accurate forecasting.

Electricity Consumption Analysis for West Region

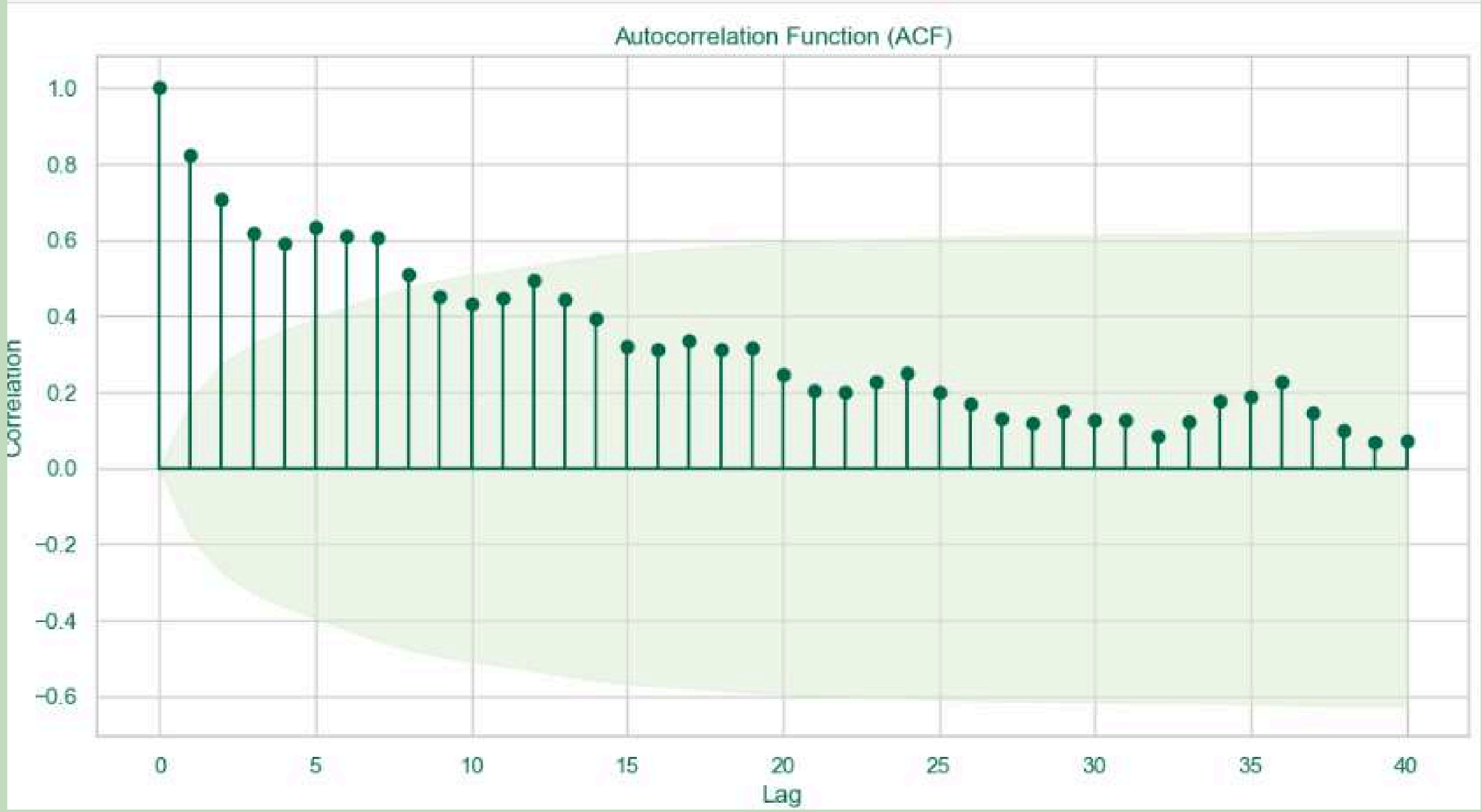
Time Series Plot



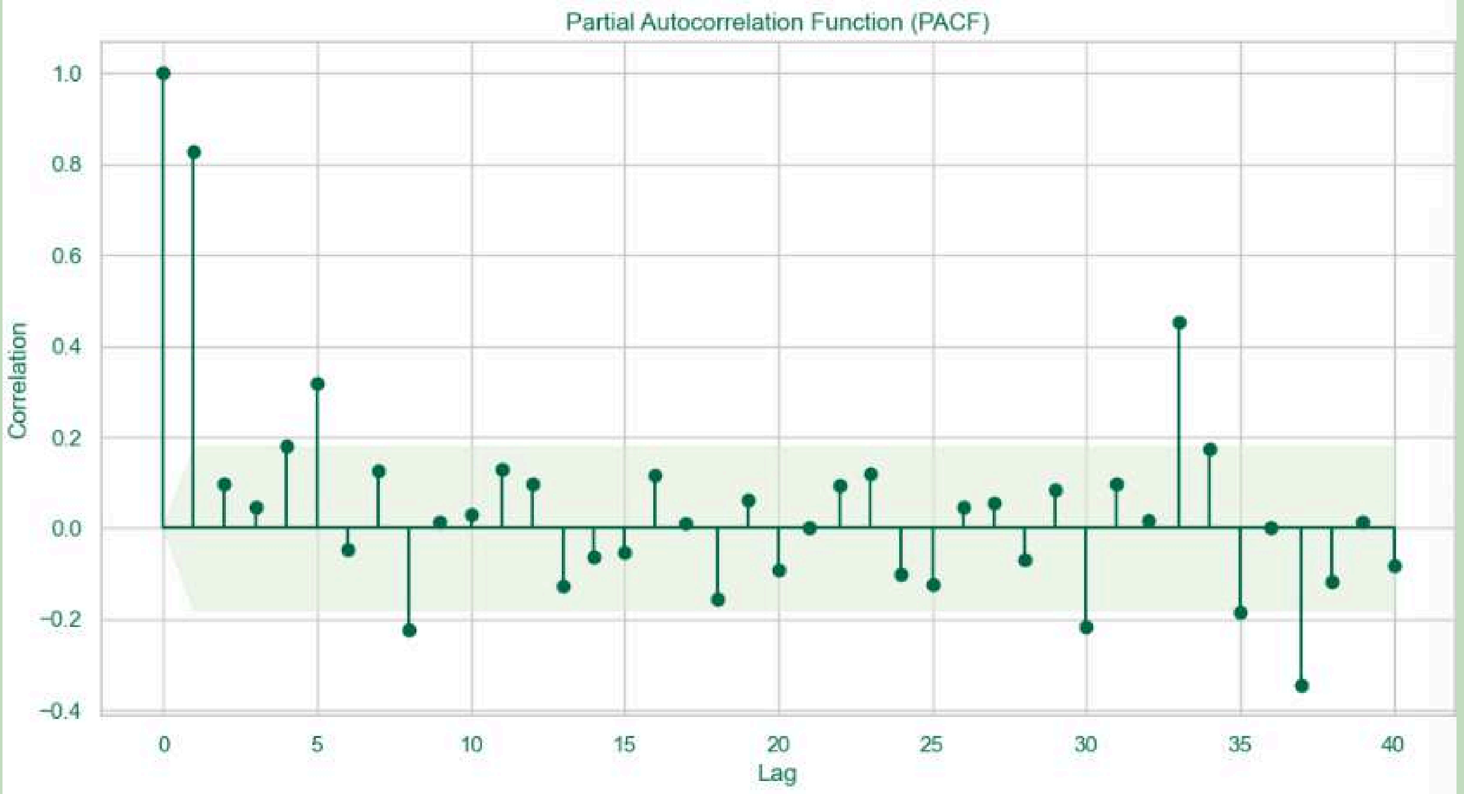
Time Series Decomposition Plot



Autocorrelation Function (ACF) Plot



Partial Autocorrelation Function (PACF) Plot



Stationarity Check

Augmented Dickey-Fuller (ADF) Test Results

```
result = adfuller(train_data)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -0.150907
p-value: 0.944126
```

- Negative ADF Statistic
- A p-value of 0.944126 (which is greater than L.o.S 0.05)

Thus, we fail to reject the Null Hypothesis of the data being non-stationary and can safely conclude that our data is non-stationary.

Also, from the decomposition performed, we can also conclude that seasonality is present in the data

Preprocessing for Stationarity: Seasonal and Normal Differencing

After Seasonal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
sd_df_train = train_data - train_data.shift(12)
sd_df_train = sd_df_train.dropna()
result = adfuller(sd_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -2.391947
p-value: 0.144022
```

After Normal Differencing

Augmented Dickey-Fuller (ADF) Test Results

```
double_df_train= sd_df_train - sd_df_train.shift(1)
double_df_train= double_df_train.dropna()
result = adfuller(double_df_train)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

```
ADF Statistic: -4.258358
p-value: 0.000524
```

After applying normal differencing, it shows stationarity with a p-value below 0.05, affirming the stationary nature of the data.

Holt-Winters Linear Exponential Smoothing Model Results

Holt-Winters Model Summary

ExponentialSmoothing Model Results

Dep. Variable:	West	**No. Observations:**	120
Model:	ExponentialSmoothing	**SSE**	430489699.655
Optimized:	True	**AIC**	1843.153
Trend:	Additive	**BIC**	1887.753
Seasonal:	Additive	**AICC**	1849.925
Seasonal Periods:	12	**Date:**	Wed, 04 Sep 2024
Box-Cox:	False	**Time:**	06:37:16
Box-Cox Coeff.:	None		

	coeff	code	optimized
smoothing_level	0.6379587	alpha	True
smoothing_trend	0.0035501	beta	True
smoothing_seasonal	0.0020589	gamma	True

Holt-Winters Model Equations

Holt-Winters Model

The Holt-Winters model equations based on the provided model summary are as follows:

Level Equation

$$L_t = 0.6379587(Y_t - S_t) + (1 - 0.6379587)(L_{t-1} + T_{t-1})$$

Trend Equation

$$T_t = 0.0035501(L_t - L_{t-1}) + (1 - 0.0035501)T_{t-1}$$

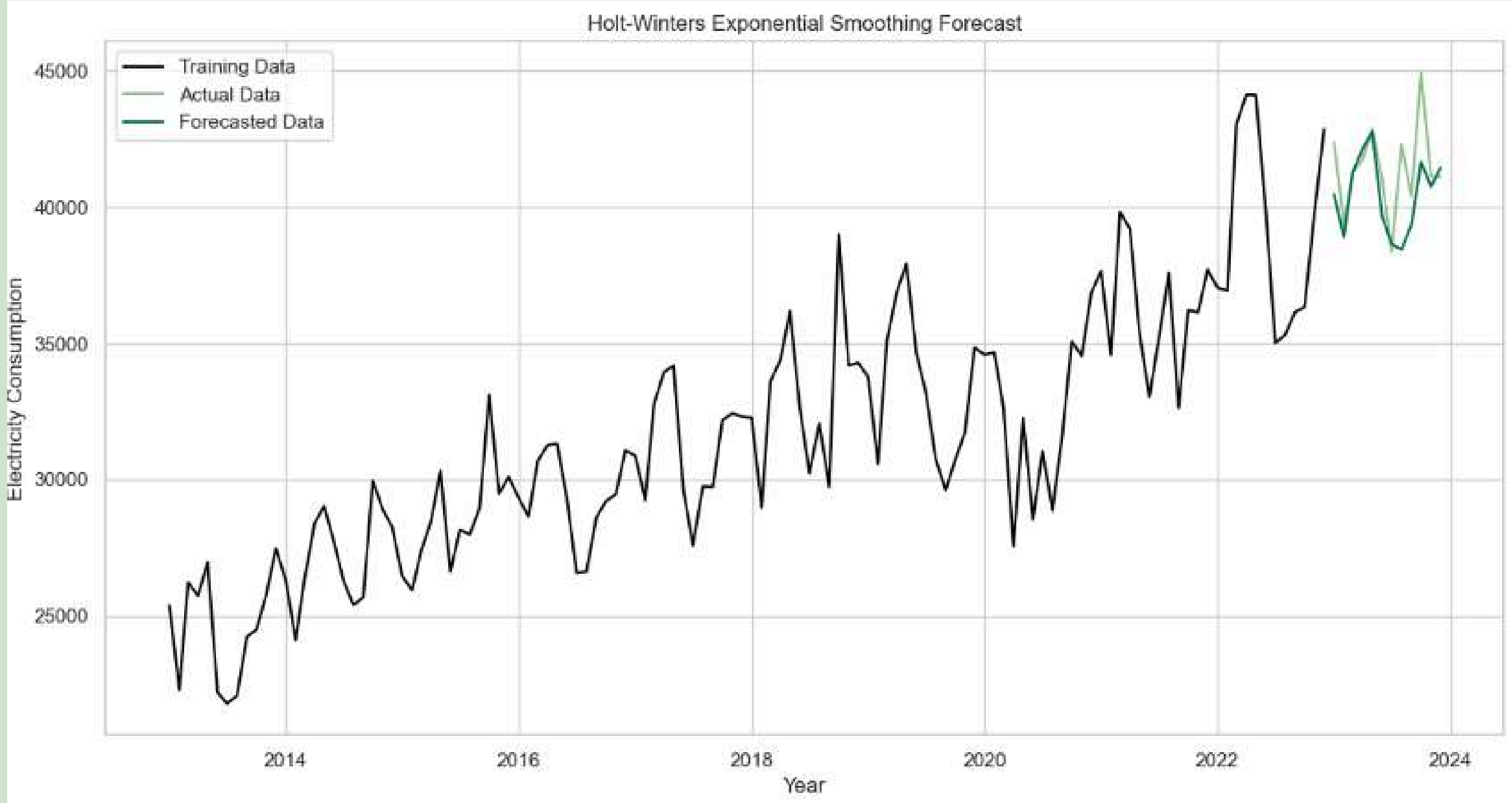
Seasonal Component

$$S_t = 0.0020589(Y_t - L_t) + (1 - 0.0020589)S_{t-12}$$

Forecast Equation

$$\hat{Y}_{t+h} = L_t + h \cdot T_t + S_{t+12-h}$$

Holt-Winters Model Fit and Forecasts



Holt-Winters Model Accuracy Metrics

MAPE
(Mean Absolute Percentage Error) → **2.65%**

MSE
(Mean Squared Error) → **2743653.148060531**

SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Summary

Dep. Variable:	West	No. Observations:	120
Model:	SARIMAX(0, 1, 0)x(1, 1, [1, 2], 12)	Log Likelihood	-749.900
Date:	Wed, 04 Sep 2024	AIC	1507.801
Time:	06:51:54	BIC	1517.427
Sample:	01-01-2013	HQIC	1511.666
	- 12-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.S.L12	0.1659	0.228	0.727	0.467	-0.281	0.613
ma.S.L12	-1.3301	0.245	-5.428	0.000	-1.810	-0.850
ma.S.L24	0.4779	0.185	2.586	0.010	0.116	0.840
sigma2	4.699e+06	7.26e+05	6.475	0.000	3.28e+06	6.12e+06

Ljung-Box (L1) (Q):	2.27	Jarque-Bera (JB):	0.66
Prob(Q):	0.13	Prob(JB):	0.72
Heteroskedasticity (H):	3.76	Skew:	-0.01
Prob(H) (two-sided):	0.00	Kurtosis:	3.44

SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\text{SARIMA}(0,1,0)(1,1,1)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 1.3301 Y_{t-12} + 0.4799 Y_{t-24}$$

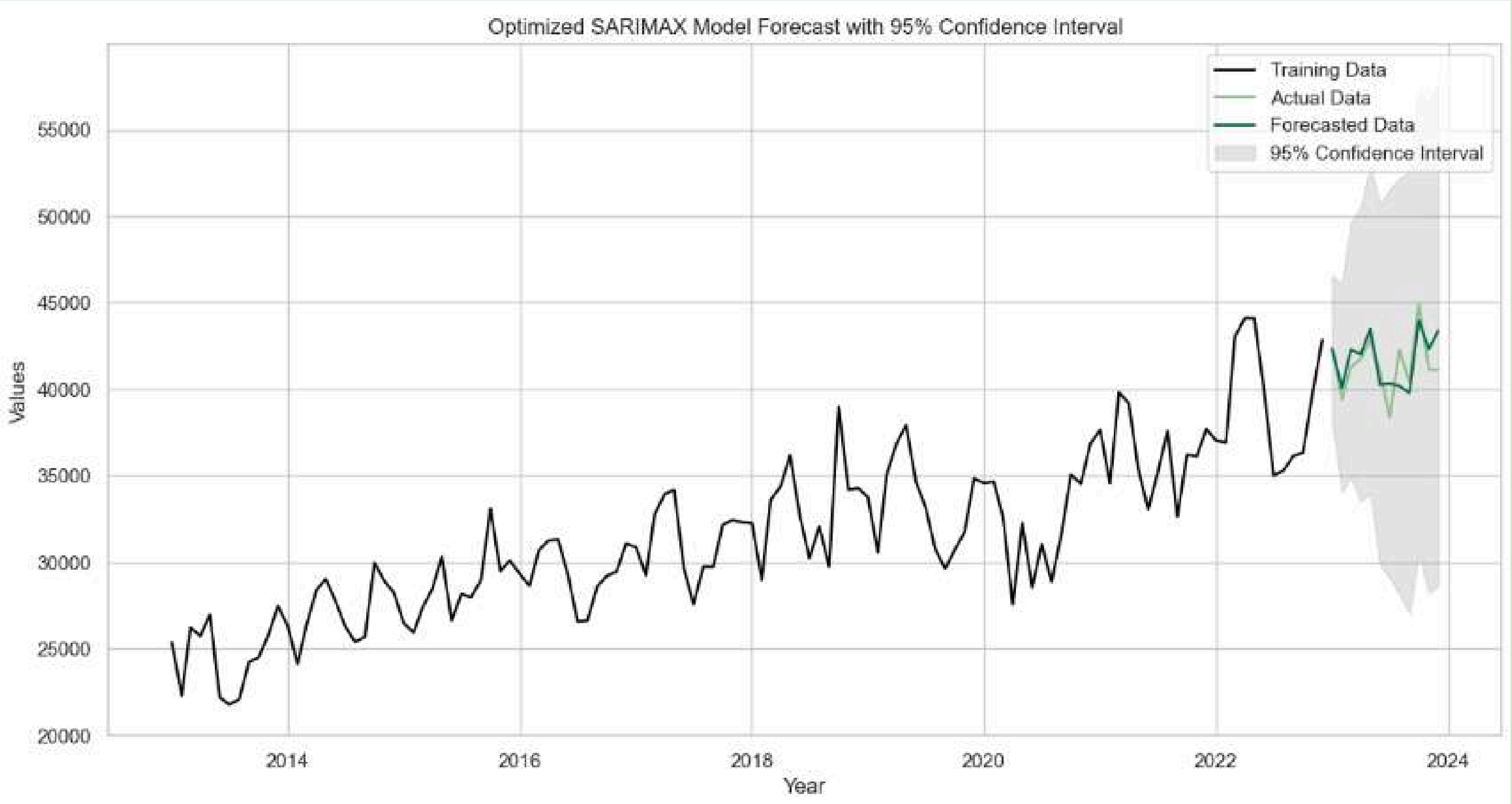
where:

- Y_{t-12} and Y_{t-24} represent the actual values at time $t-12$ and $t-24$, respectively.
- μ is the constant term (if applicable).

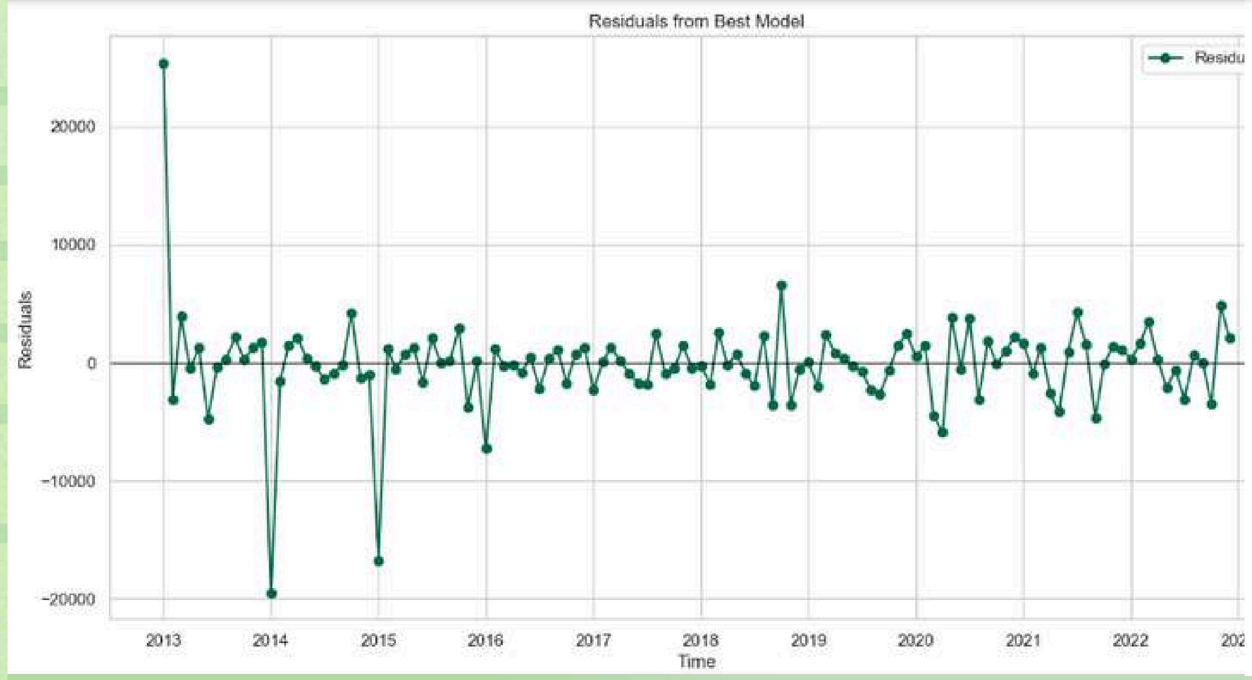
SARIMA Model Accuracy Metrics

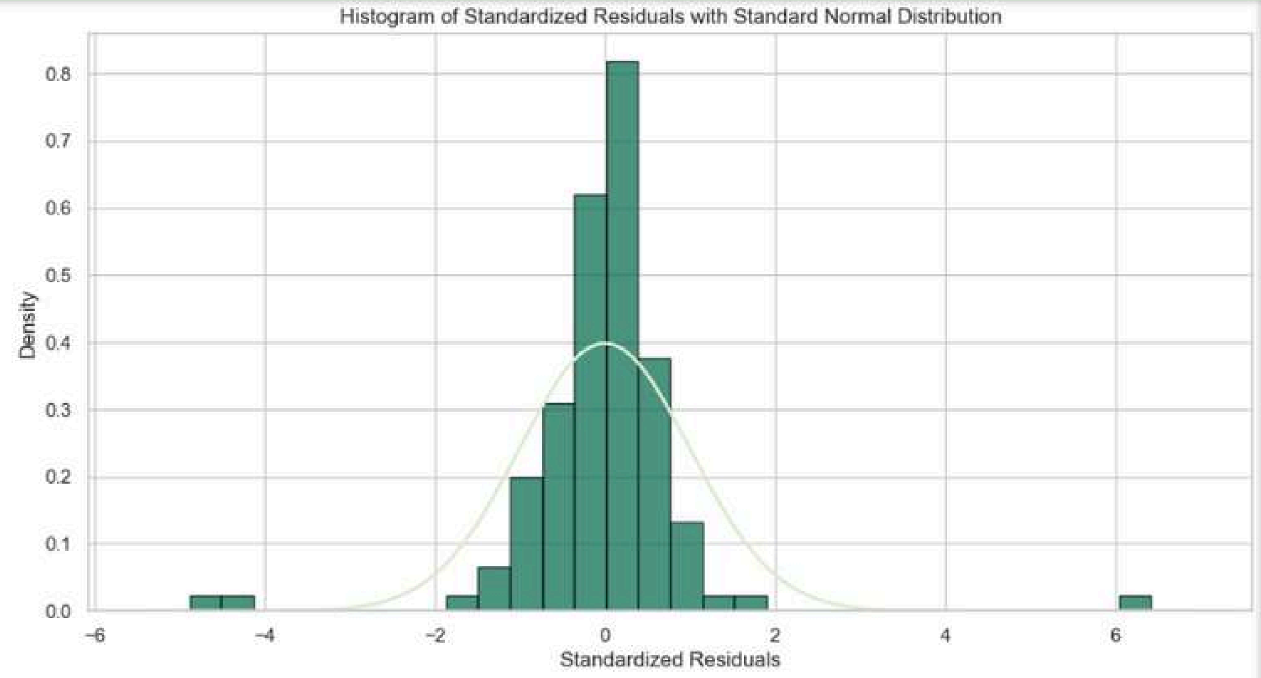
Metric	Value
MAPE (Mean Absolute Percentage Error)	**2.54%**
MSE (Mean Squared Error)	**1548040.812904063**

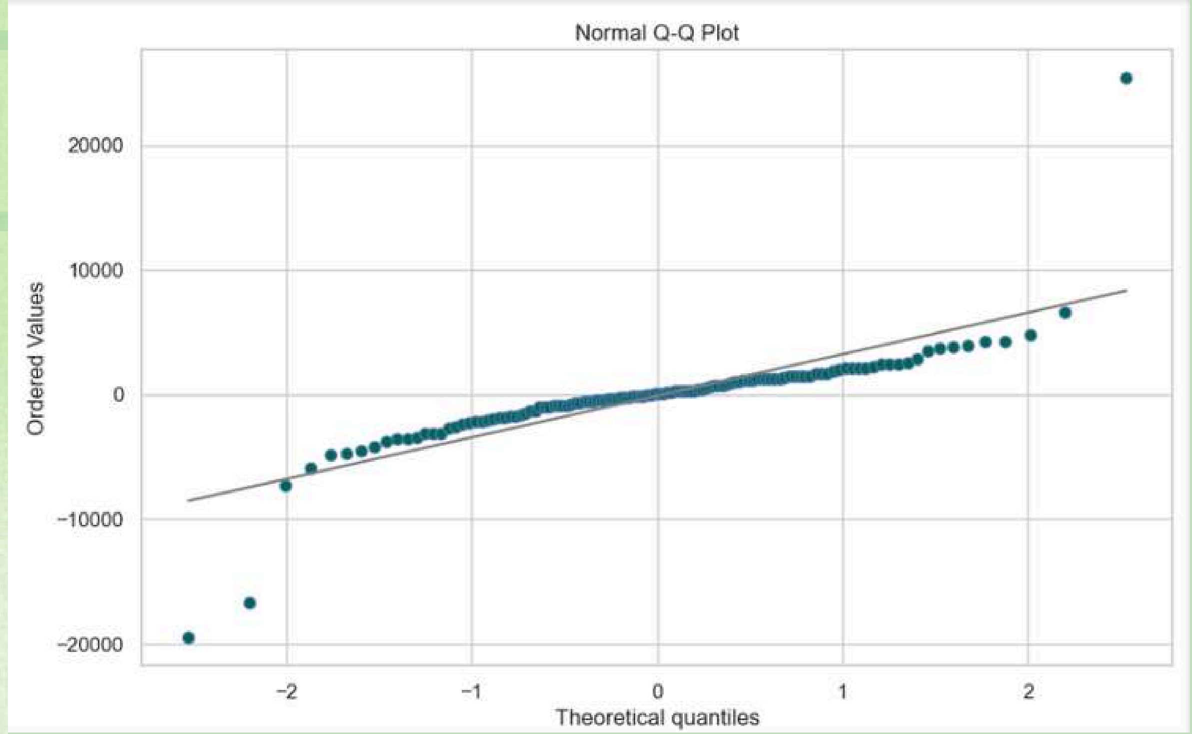
SARIMA Model Fit and Forecasts

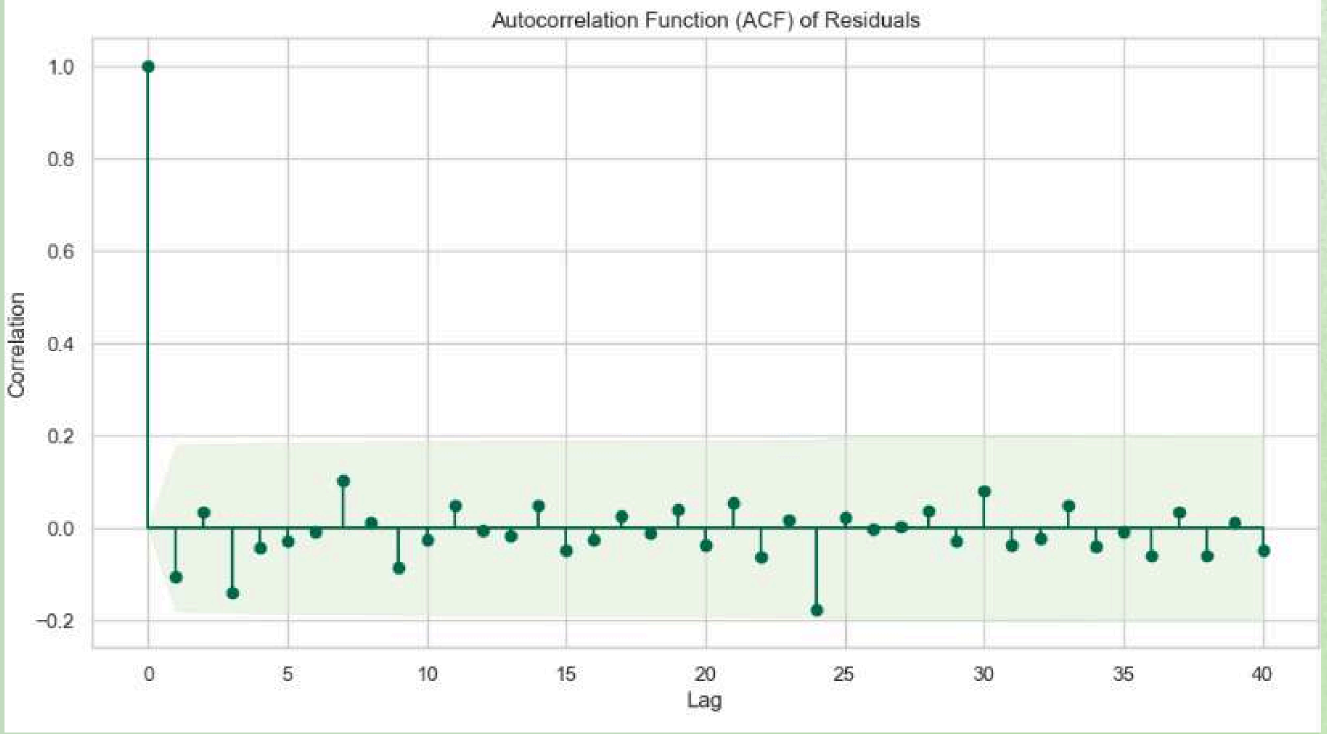


Residual Analysis









The residual analysis confirms that the SARIMA model satisfies all key assumptions:

- **Normality**: Residuals are normally distributed.
- **Stationarity**: No trends or patterns in the residuals.
- **No Autocorrelation**: Residuals show no significant autocorrelation.
- **Homoscedasticity**: Residuals have constant variance.

This indicates a good model fit for accurate forecasting.

Forecast for Top 3 States

To gain deeper insights into regional electricity consumption patterns, we conducted individual time series forecasts for the top three states with the highest consumption. This analysis allows us to capture state-level trends and seasonality, providing more targeted and granular forecasts that complement the regional predictions.

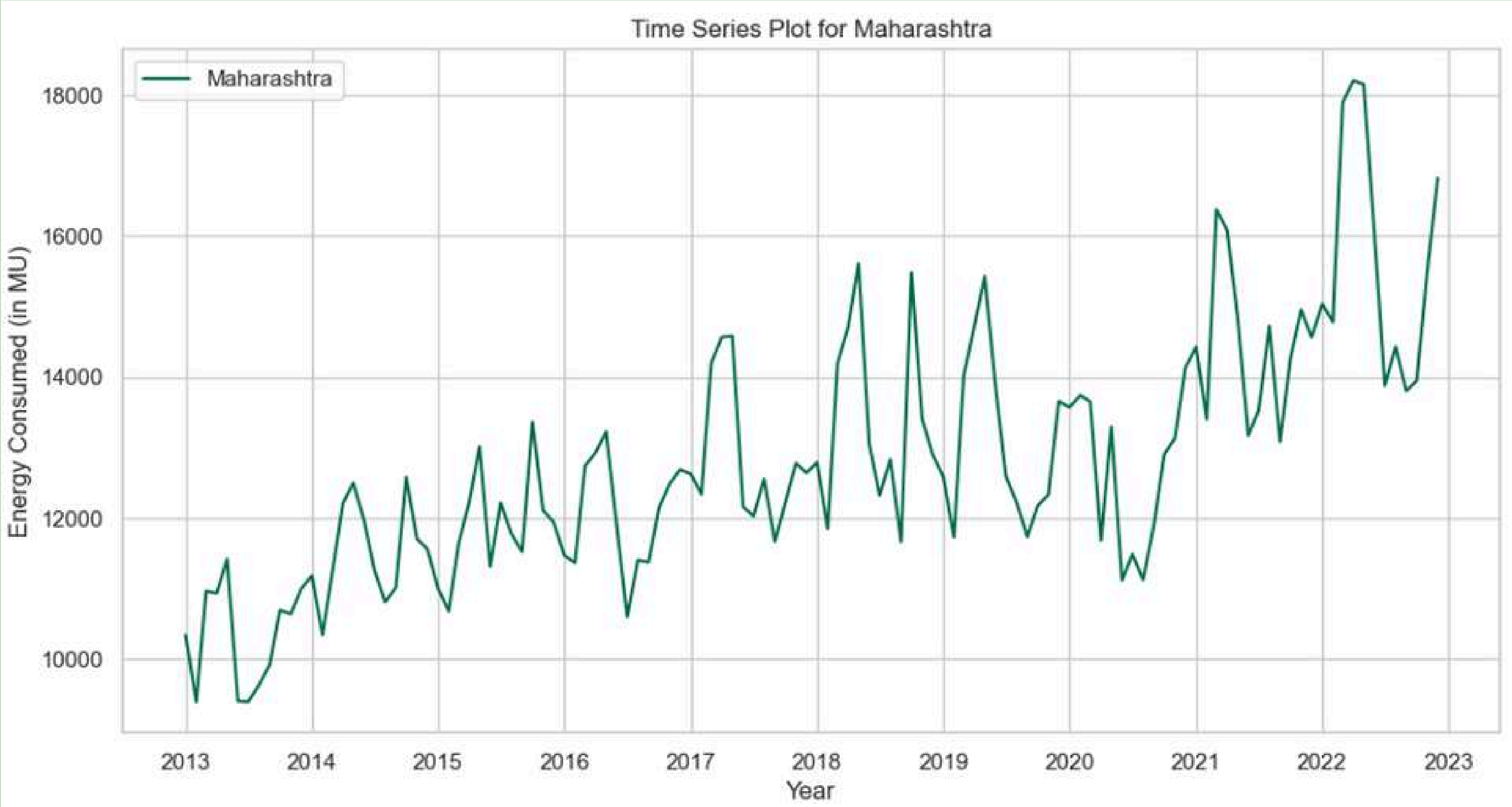
01.
Maharashtra
(1735108.8 MU)

02.
Gujarat
(1260955.3 MU)

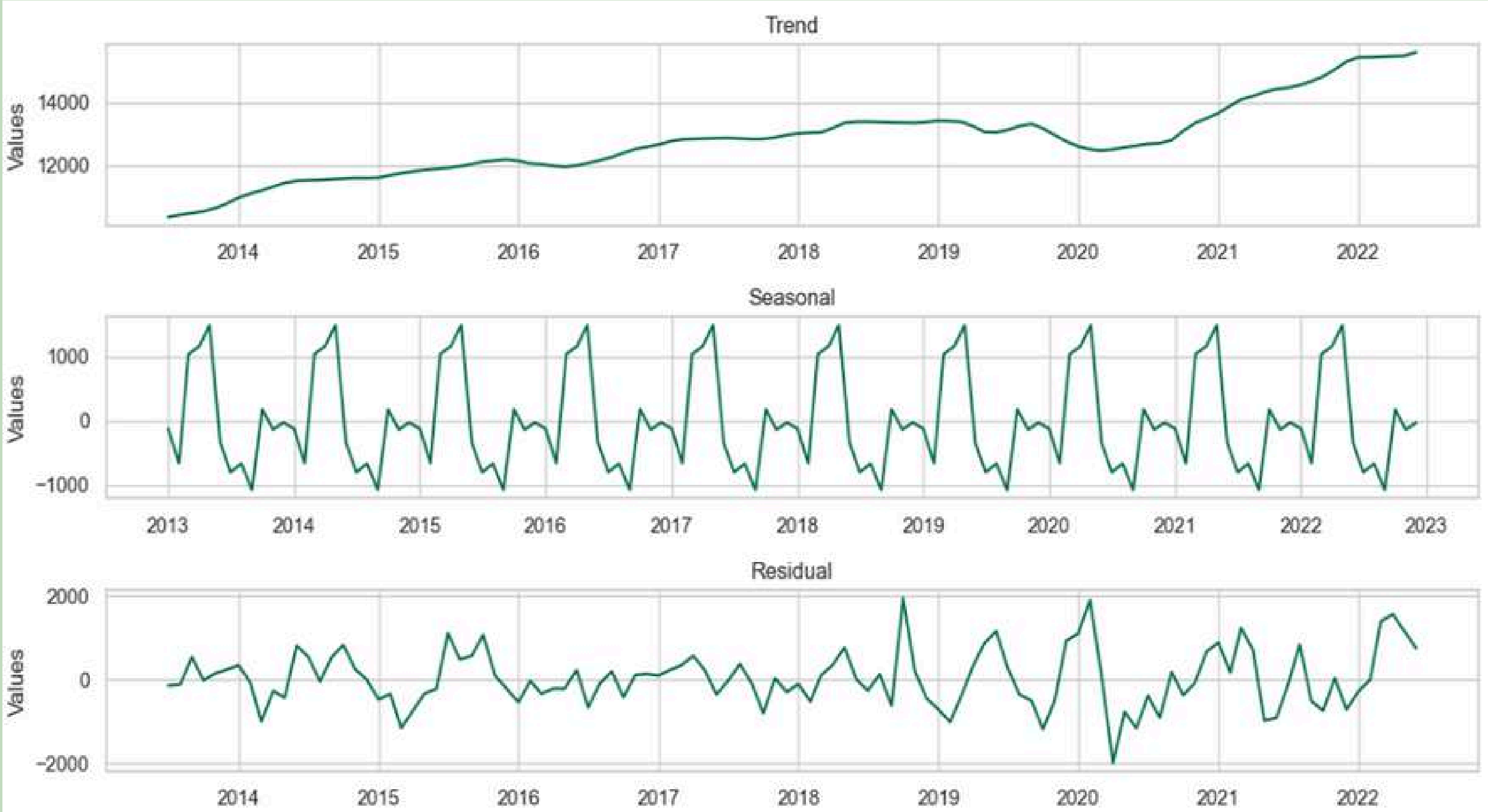
03.
Uttar Pradesh
(1250038.8 MU)

01. Maharashtra

Time Series Plot

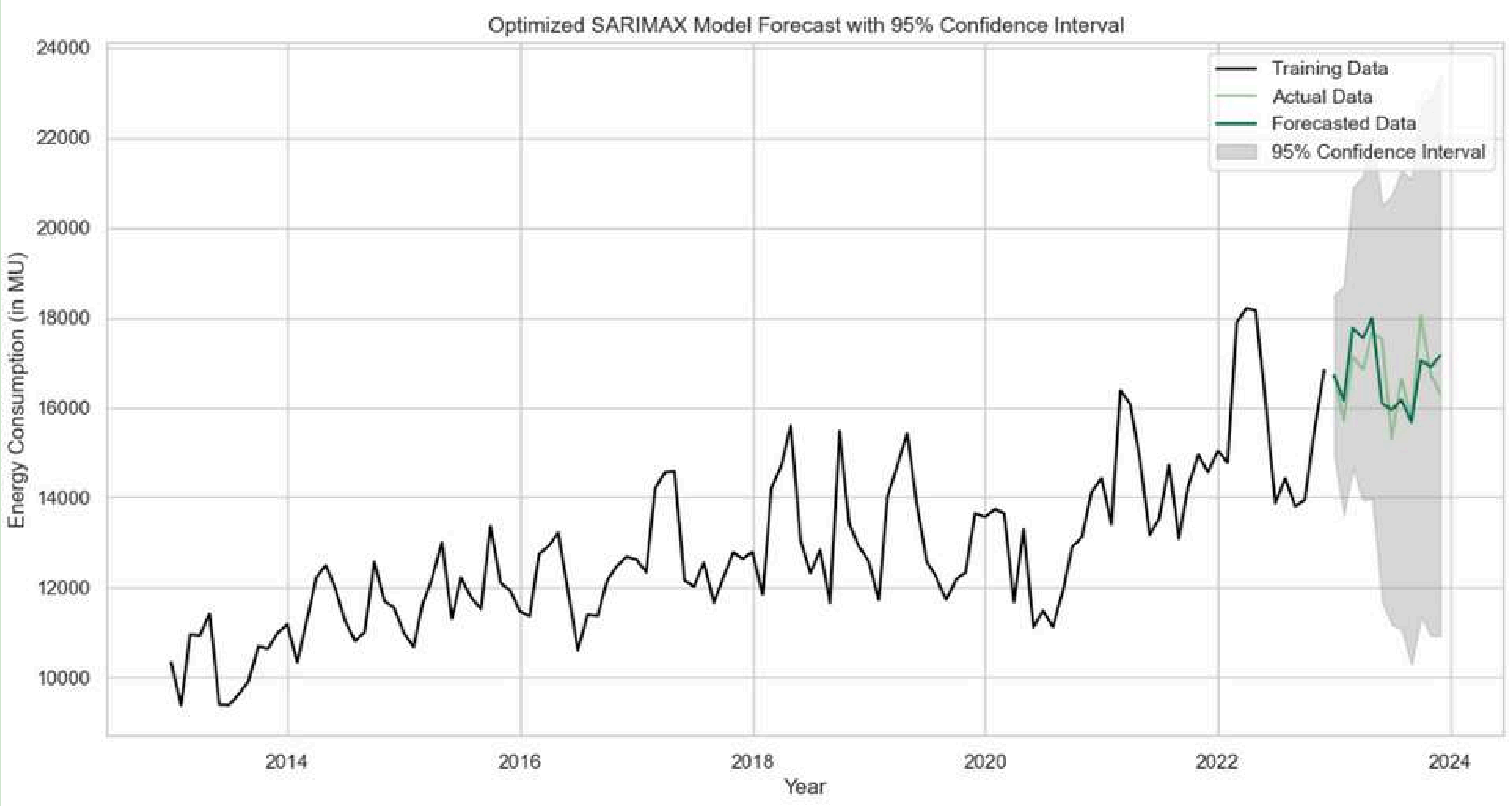


Time Series Decomposition Plot



SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Fit and Forecasts



SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\mathrm{SARIMA}(0,1,0)(1,1,1)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.0338Y_{t-12} - 3.4137\epsilon_{t-12} + 1.8039\epsilon_{t-24}$$

where:

- Y_{t-12} represents the actual value at time $t-12$.
- ϵ_{t-12} and ϵ_{t-24} are the error terms at time $t-12$ and $t-24$, respectively.
- μ is the constant term (if applicable).

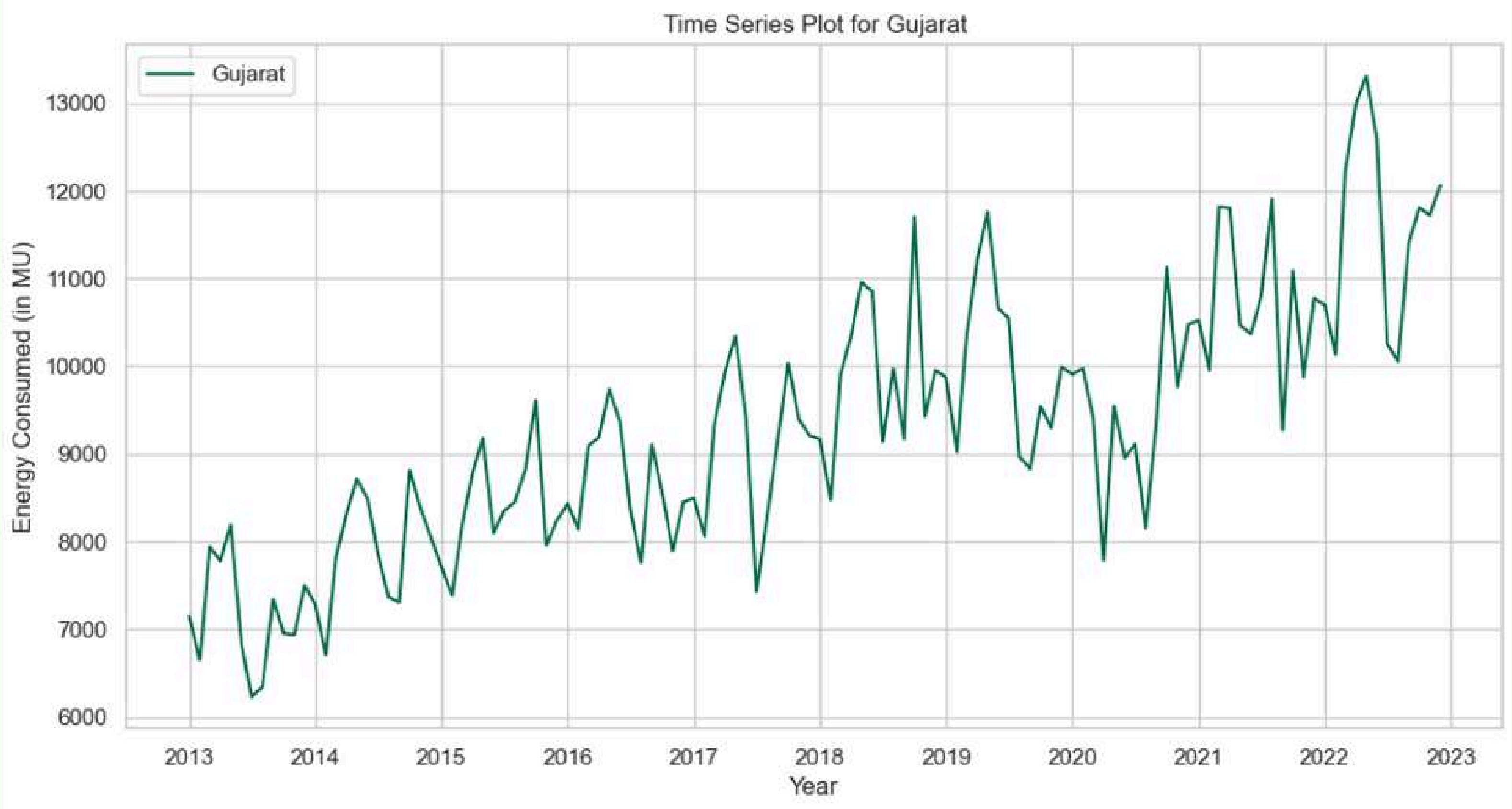
SARIMA Model Accuracy Metrics

MAPE
(Mean Absolute Percentage Error) → **3.35%**

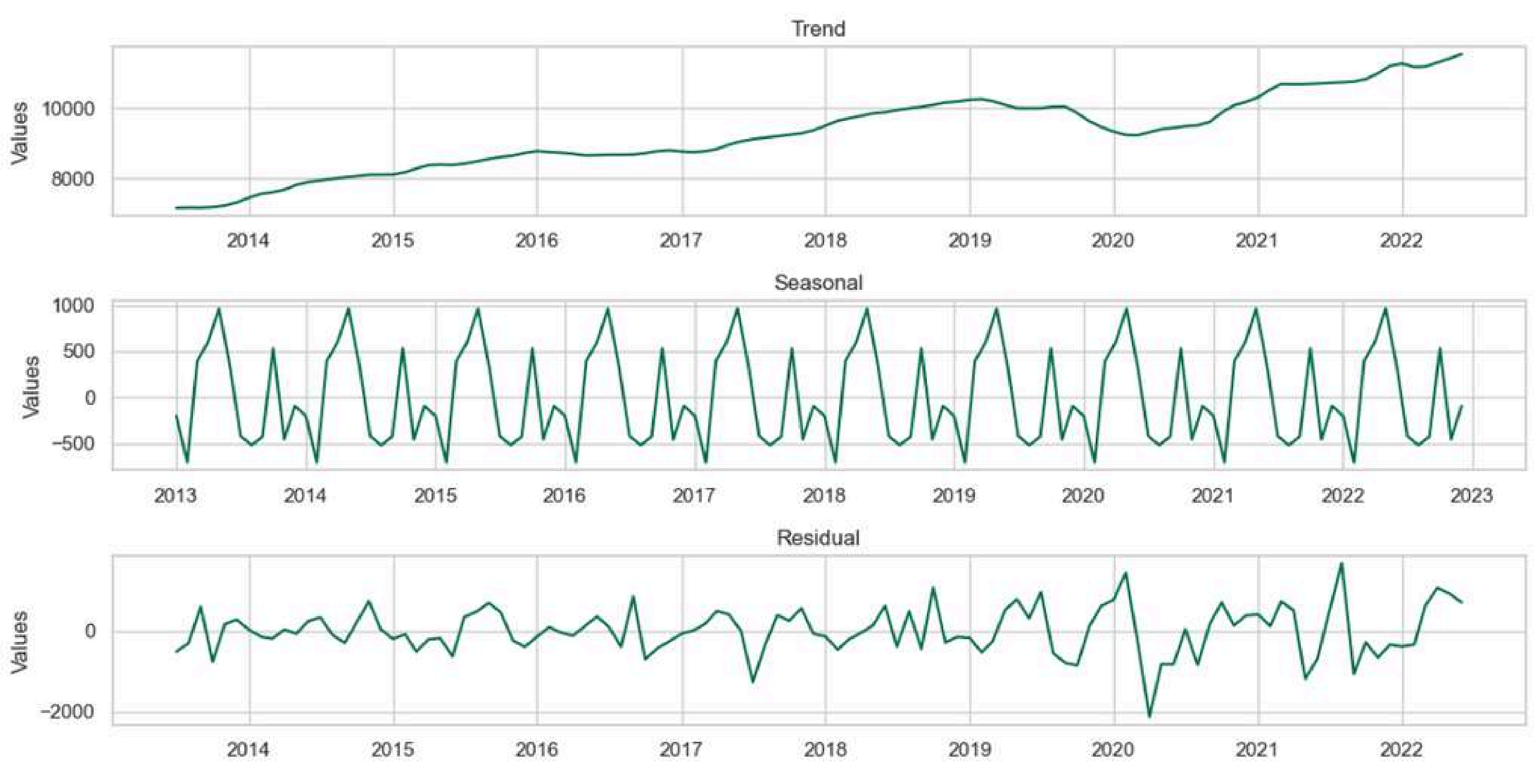
MSE
(Mean Squared Error) → **469982.54473243834**

02. Gujarat

Time Series Plot

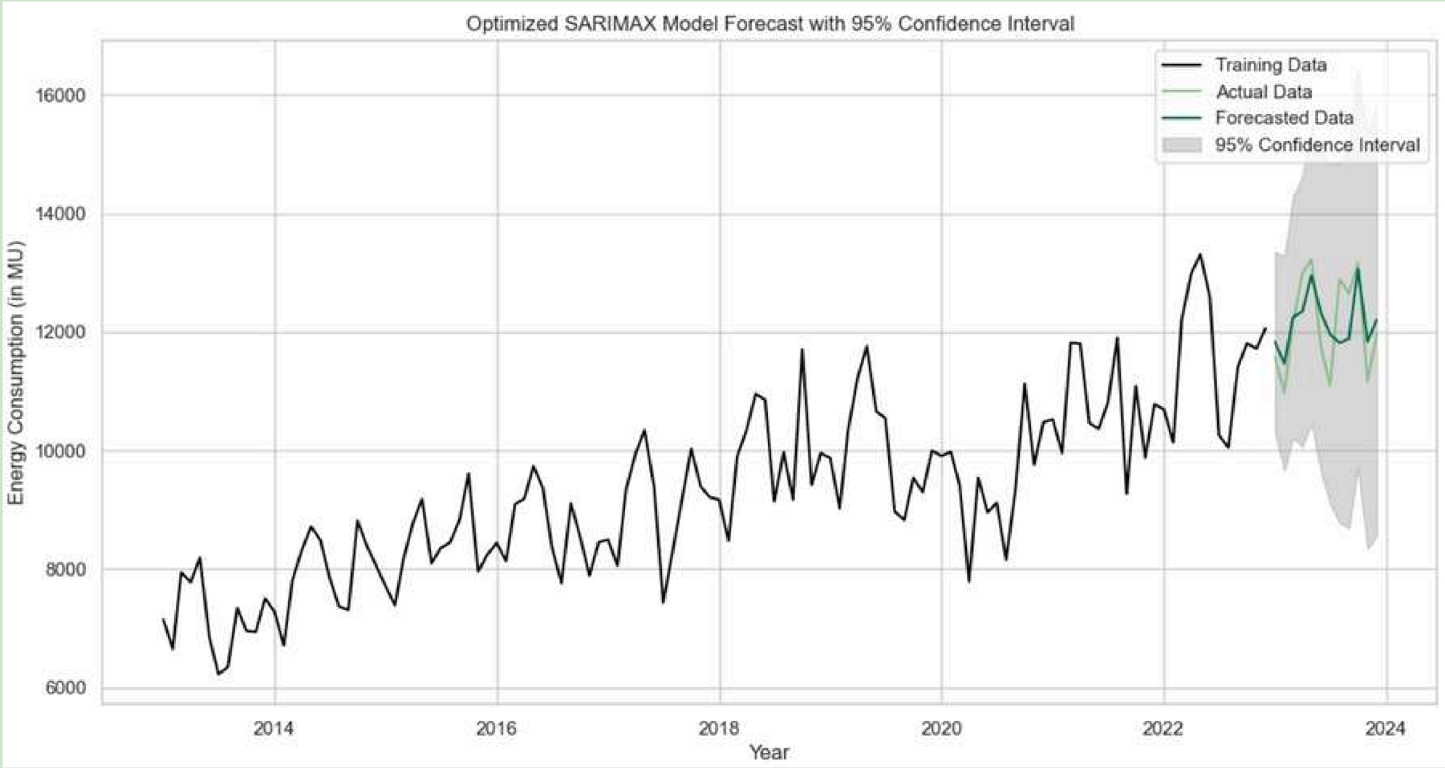


Time Series Decomposition Plot



SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Fit and Forecasts



SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\text{SARIMA}(2,1,0)(0,1,1)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.3499Y_t - 0.1603Y_{t-1} - 1.3342\epsilon_{t-12} + 0.3334\epsilon_{t-24}$$

where:

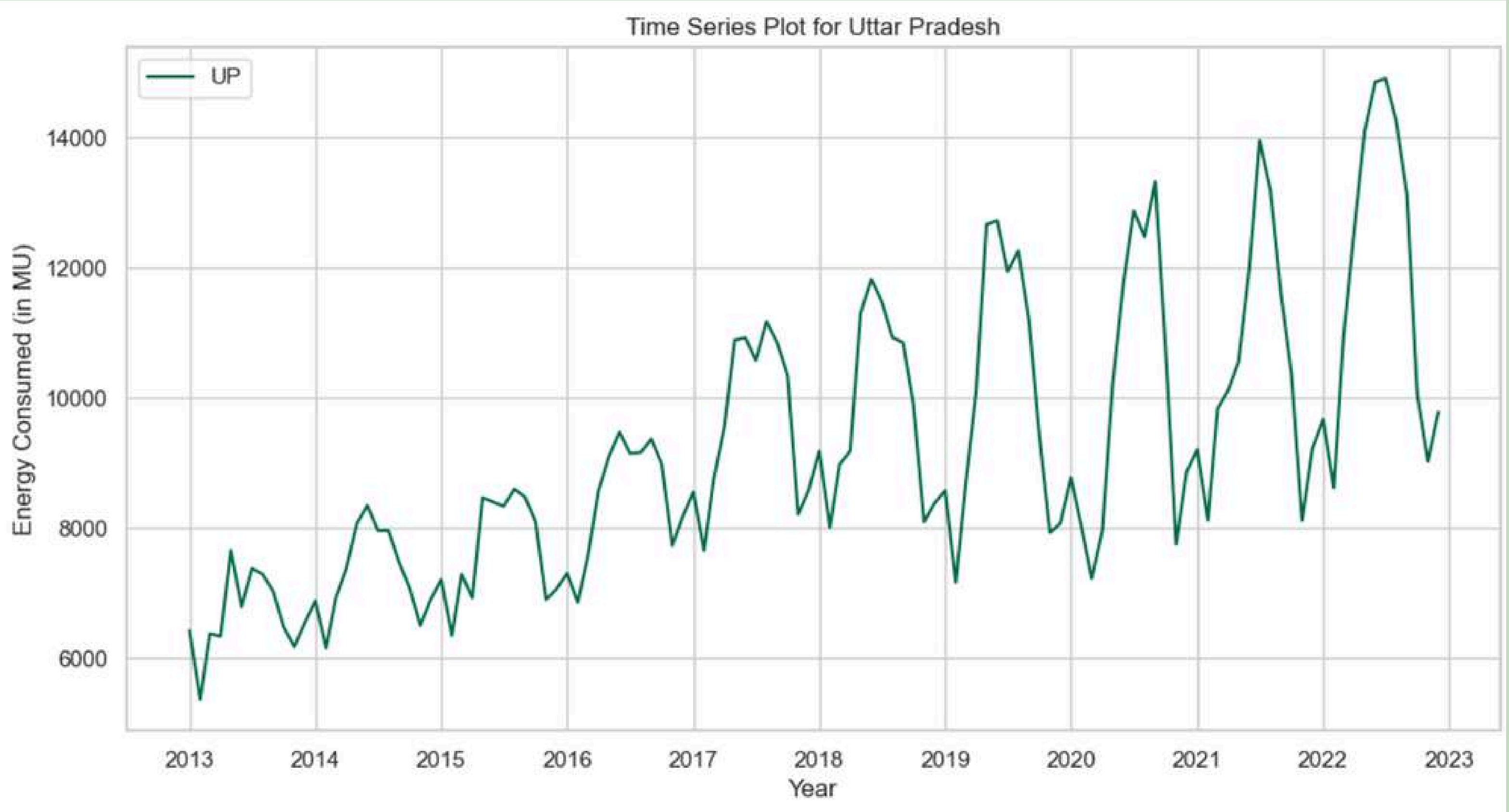
- Y_t and Y_{t-1} represent the actual values at time t and $t-1$, respectively.
- ϵ_{t-12} and ϵ_{t-24} are the seasonal error terms at time $t-12$ and $t-24$, respectively.
- μ is the constant term (if applicable).

SARIMA Model Accuracy Metrics

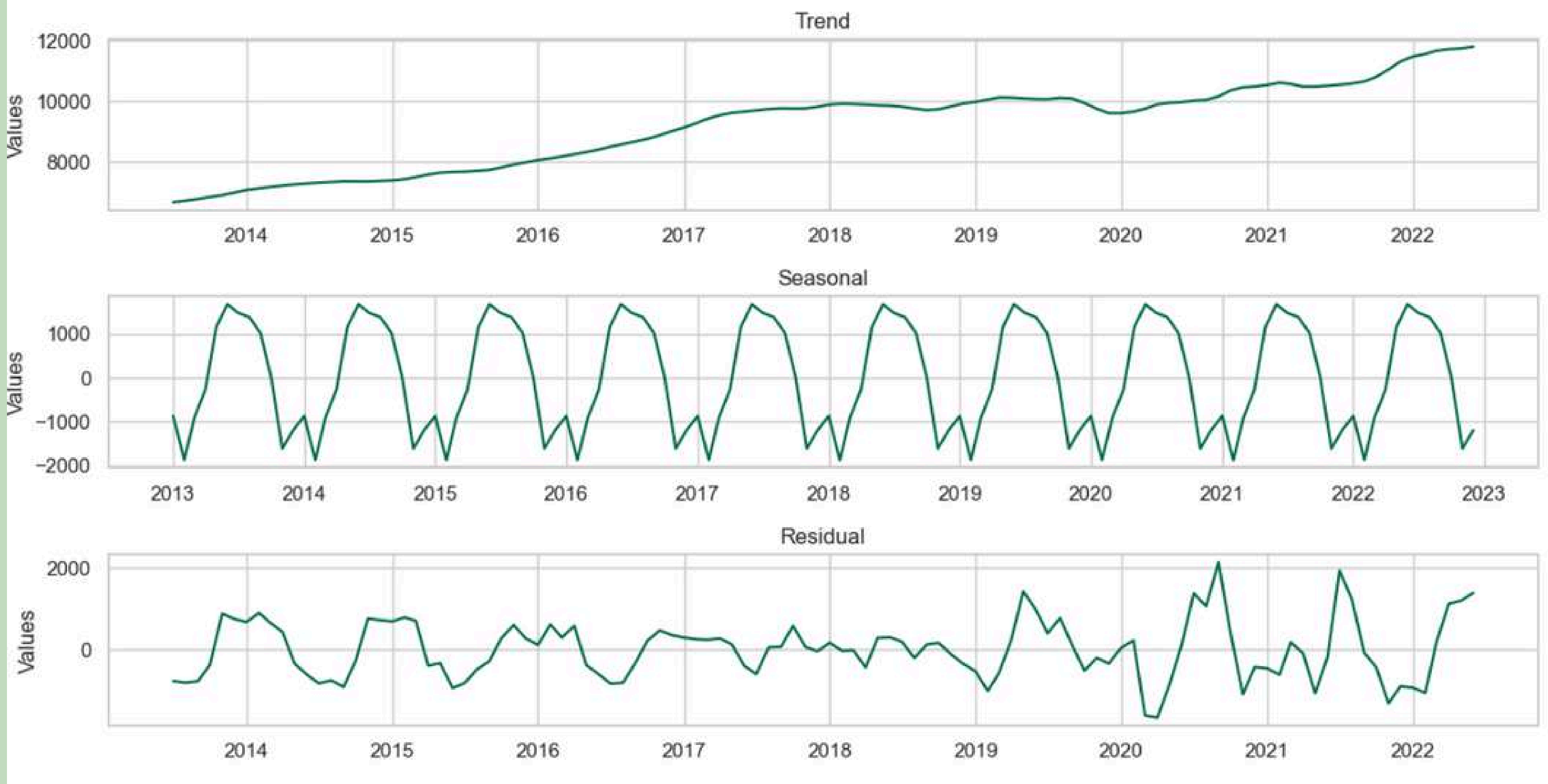
MAPE (Mean Absolute Percentage Error)	**4.21%**
MSE (Mean Squared Error)	**345068.12192847626**

03. Uttar Pradesh

Time Series Plot

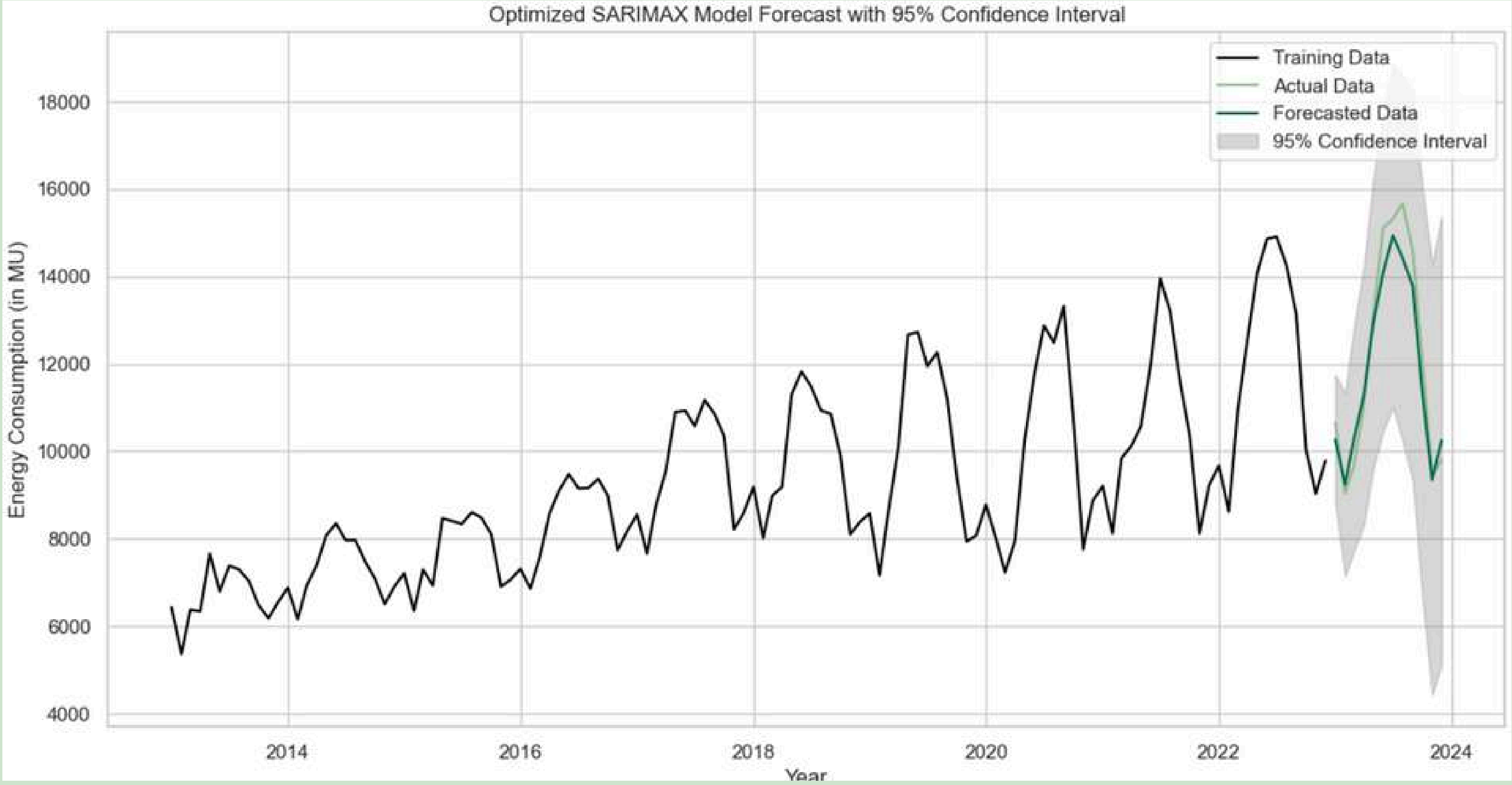


Time Series Decomposition Plot



SARIMA (Seasonal Autoregressive Integrated Moving Average) Model

SARIMA Model Fit and Forecasts



SARIMA Model Equation

Model Representation

The SARIMA model can be represented as:

$$\mathrm{SARIMA}(0,1,0)(2,1,1)_{12}$$

Forecast Equation

The forecast for $\hat{Y}_{t+1}$ based on the SARIMA model is given by:

$$\hat{Y}_{t+1} = \mu - 0.0699Y_{t-12} - 0.2703Y_{t-24} - 0.6147\epsilon_{t-12} + 0.2712\epsilon_{t-24}$$

where:

- Y_{t-12} and Y_{t-24} represent the actual values at time $t-12$ and $t-24$, respectively.
- ϵ_{t-12} and ϵ_{t-24} are the error terms at time $t-12$ and $t-24$, respectively.
- μ is the constant term (if applicable).

SARIMA Model Accuracy Metrics

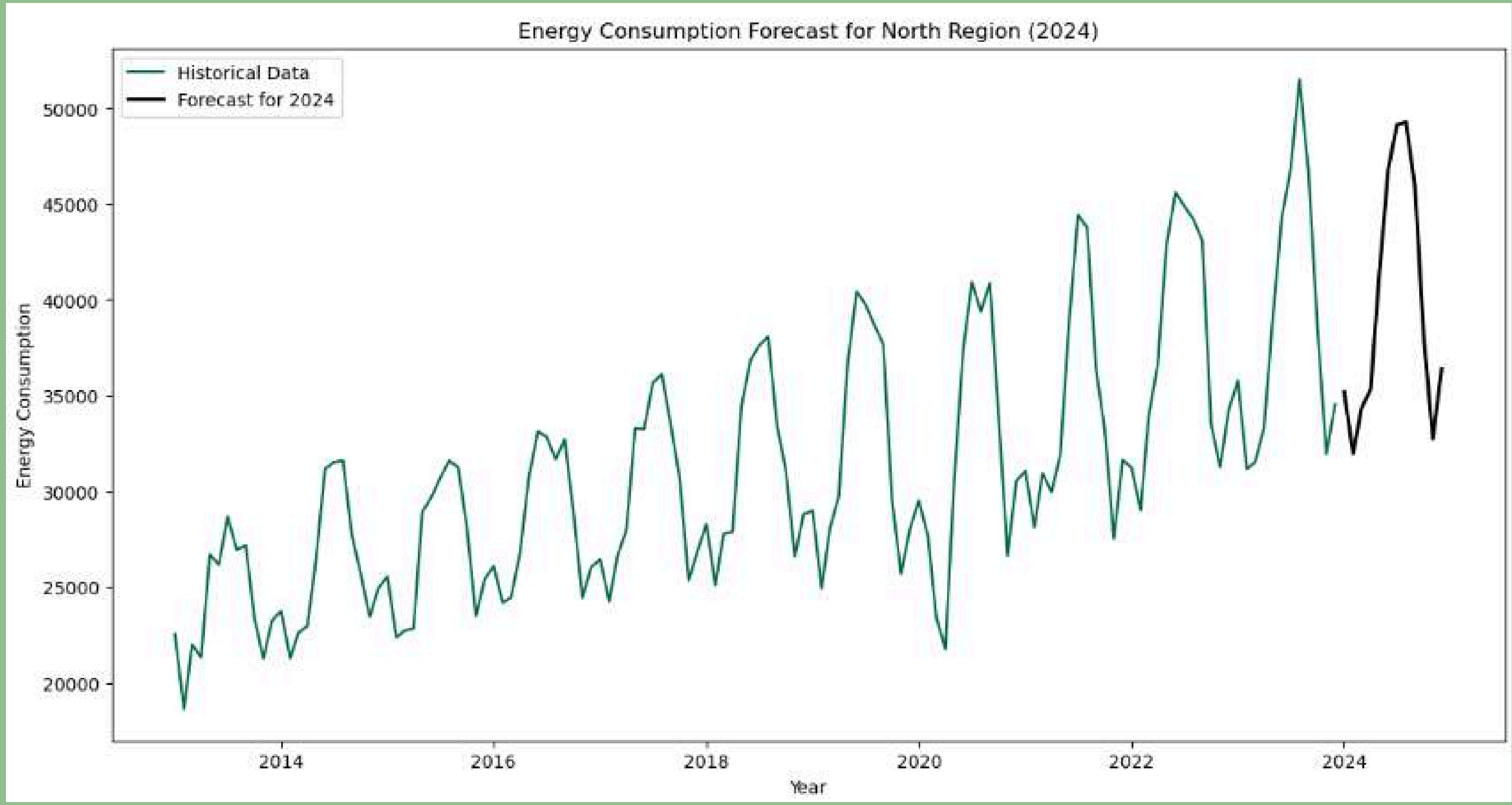
Metric	Value
MAPE (Mean Absolute Percentage Error)	**4.20%**
MSE (Mean Squared Error)	**404002.114883166**

2024 Forecast: Regional Electricity Consumption

Based on the analysis conducted so far, we utilized the SARIMA models developed for each region — North, South, East, and West — to forecast electricity consumption for 2024. These models, built on historical data from 2013 to 2023, effectively capture regional trends and seasonal patterns, providing robust predictions that can support strategic planning and decision-making for the future.

Electricity Forecast for 2024: North Region

Graphical Forecast of Electricity Consumption: North Region (2024)

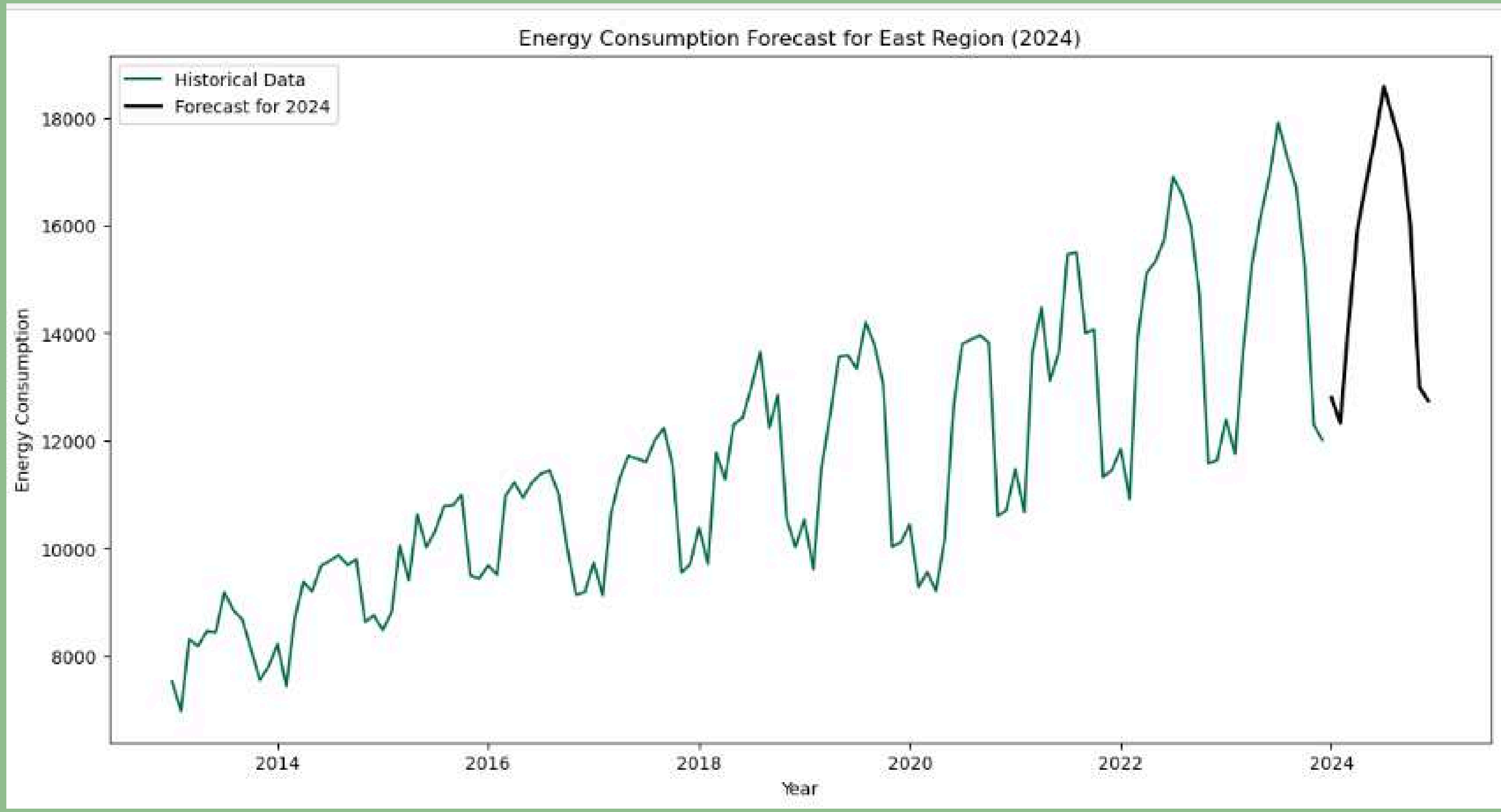


Predicted Consumption Values: North Region (2024)

Date	North
2024-01-01	35186.787160
2024-02-01	31969.782938
2024-03-01	34352.471343
2024-04-01	35313.589032
2024-05-01	41363.847003
2024-06-01	46846.168858
2024-07-01	49137.492845
2024-08-01	49267.318938
2024-09-01	45836.152952
2024-10-01	38012.889170
2024-11-01	32736.032074
2024-12-01	36393.063135

Electricity Forecast for 2024: East Region

Graphical Forecast of Electricity Consumption: East Region (2024)

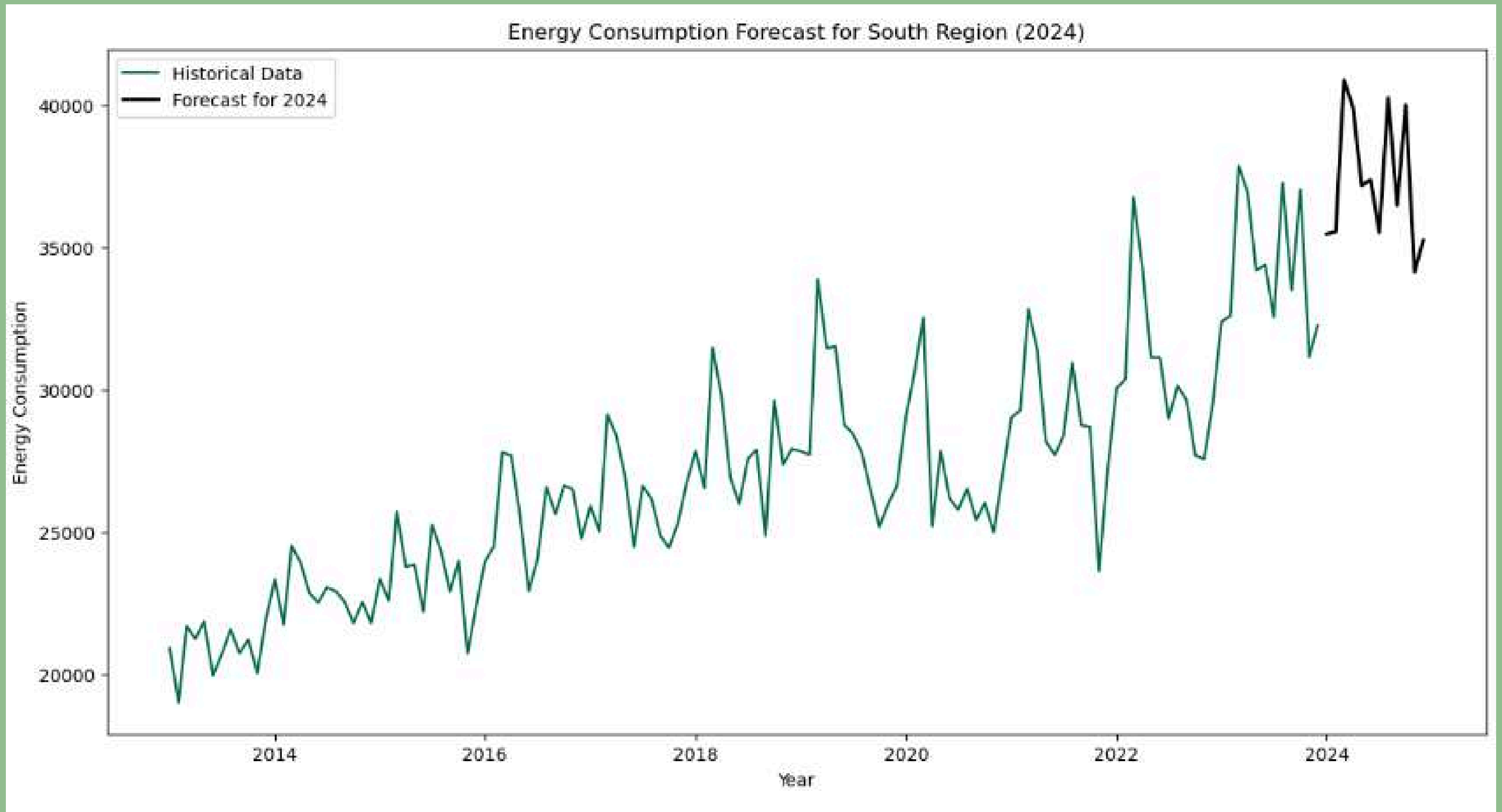


Predicted Consumption Values: East Region (2024)

Date	East
2024-01-01	12802.901959
2024-02-01	12329.833859
2024-03-01	14208.903608
2024-04-01	15945.584660
2024-05-01	16798.708491
2024-06-01	17640.027376
2024-07-01	18575.250547
2024-08-01	17991.201763
2024-09-01	17396.070797
2024-10-01	15976.002448
2024-11-01	12997.312529
2024-12-01	12742.325229

Electricity Forecast for 2024: South Region

Graphical Forecast of Electricity Consumption: South Region (2024)

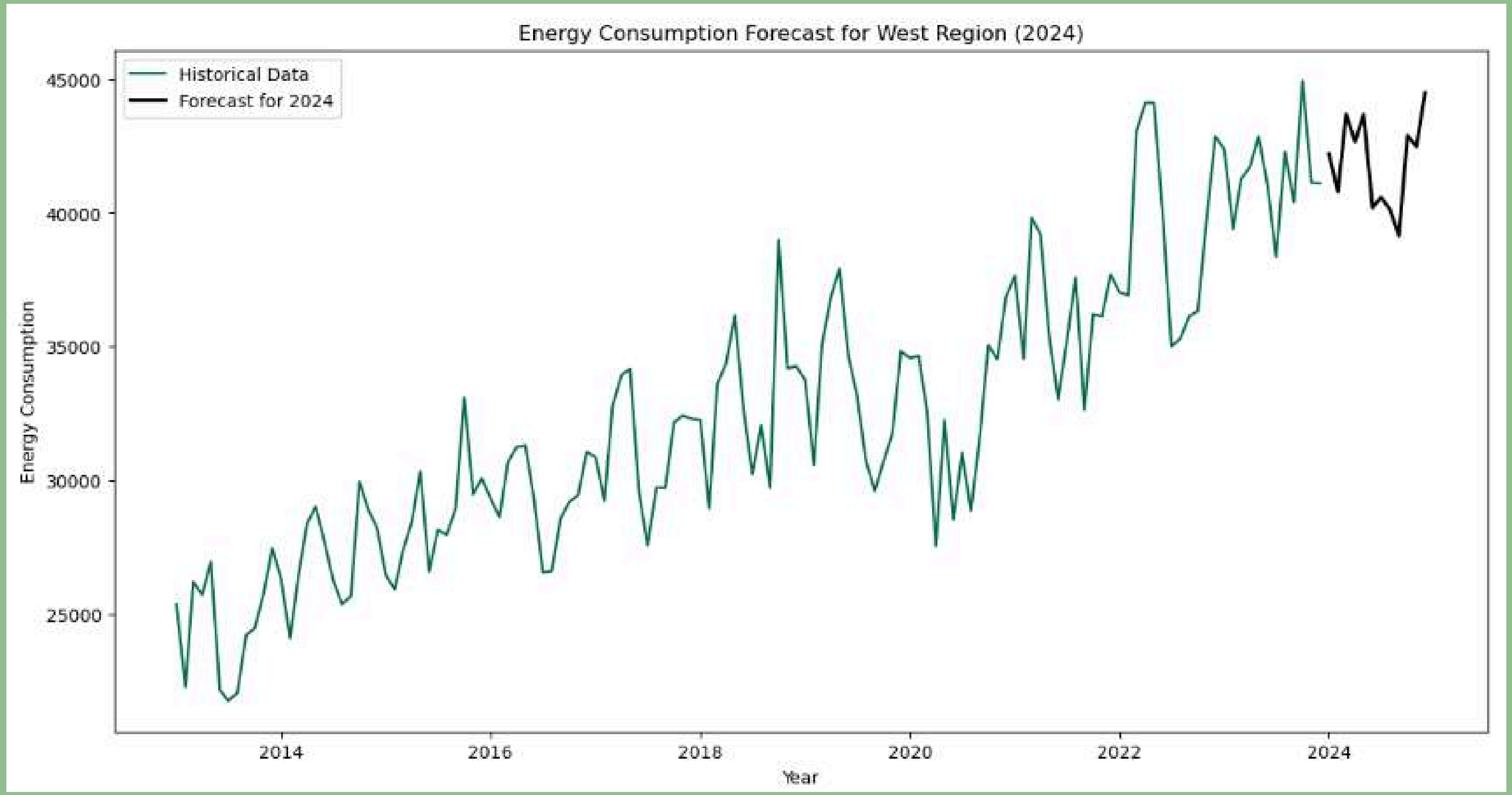


Predicted Consumption Values: South Region (2024)

Date	South
2024-01-01	35470.969254
2024-02-01	35544.688967
2024-03-01	40874.663503
2024-04-01	39922.718787
2024-05-01	37167.522520
2024-06-01	37371.469168
2024-07-01	35525.663847
2024-08-01	40249.196407
2024-09-01	36479.379302
2024-10-01	40001.307741
2024-11-01	34130.335740
2024-12-01	35248.324785

Electricity Forecast for 2024: West Region

Graphical Forecast of Electricity Consumption: West Region (2024)



Predicted Consumption Values: West Region (2024)

Date	West
2024-01-01	42204.695766
2024-02-01	40784.633396
2024-03-01	43685.696614
2024-04-01	42651.182788
2024-05-01	43682.147635
2024-06-01	40192.772832
2024-07-01	40586.279809
2024-08-01	40116.152846
2024-09-01	39144.974657
2024-10-01	42882.926509
2024-11-01	42475.421261
2024-12-01	44489.792230

Conclusion

The SARIMA models applied to each region's electricity consumption data have provided detailed forecasts for 2024. The forecasts reveal region-specific trends and anticipated consumption patterns, offering valuable insights for strategic planning. The North, South, East, and West regions each exhibit unique patterns, with projections reflecting both seasonal effects and underlying trends. These forecasts will support informed decision-making and resource allocation, helping to address regional demand variations effectively. Ongoing monitoring and model refinement will be essential to ensure accuracy and adapt to any unforeseen changes in consumption patterns.

THANK YOU!

Presented by Group 7